Cauchy prior for Bayesian linearized seismic AVO inversion

Odd Kolbjørnsen

Department of mathematical sciences,
Norwegian university of science and technology
Trondheim, Norway

Abstract

A Bayesian approach is used to estimate material parameters of the underground. The parameters to be estimated are pressure wave velocity, shear wave velocity and density. The data analyzed are angle gathers. The underground usually have a layered structure. A stationary log Gaussian prior model is frequently used, but is not adequate to describe a layered structure. In the current approach the prior is modeled by a superposition of a Cauchy and Gaussian processes on a logarithmic scale. The Cauchy process yields a model for the layering whereas the Gaussian processes describe fluctuations within a layer. The physics of the likelihood is approximated by a linear operator between the logarithm of the material parameters and the observed seismic traces, the error structure is assumed to be Gaussian. The final estimate is optimal under the loss criterion of absolute deviation and is evaluated by Monte Carlo integration.

The current methodology is compared to a pure log Gaussian model. The material parameters observed in a well at the Sleipner Øst field is used as a test case and synthetic seismic observations are generated. Over all the velocity estimates in the current model reduces the risk by 7%, and the average length of the 90% credibility interval is reduced by 7%. In a region where the true velocity have large fluctuations, the velocity estimate in the current model improve by 14% and 10%. In a region where the true velocity is slowly varying, the 90% credibility interval is reduced by 10%. There are only minor effects of the model concerning the estimate of the rock density.

The current model is tested for real seismic observations collected in a marine seismic survey above the Sleipner Øst field. The inversion results are satisfactory, but the information content in the observations is small due to large errors in the data.

KEY WORDS: *Bayesian statistic, Independently scattered random measures, Deconvolution, Seismic inversion*

1 Introduction

The objective of seismic inversion, is to estimate material parameters of the underground. The observations are obtained by generating an acoustic wave above the target area and record the signal reflected from the underground.

A simplified model for the wave propagation, is obtained by regarding the reflected signal as the response of a locally vertical 1D-earth model, see for example Sheriff and Geldart (1995). In this model the reflected signal can be approximated by a convolution of a wavelet and the seismic reflectivity. The seismic reflectivity is again connected to the material parameters through the Zoeppritz equation.

The problem of seismic inversion is inherently ill-posed. The high frequency components of the wave that respond to high frequency changes in the rock are dampened due to intrinsic absorption, hence an exact stable inversion is beyond reach. A stable reconstruction of the material parameters, can only be obtained by providing, directly or indirectly, information about the preferred solution. The Bayesian formalism is well suited for this task. The Bayesian choice of prior distribution is a direct way of introducing preferences in the solution space.

Stationary Gaussian random fields are frequently used to construct prior distributions, this choice is particularly successful for solving linear inverse problems, due to the simplicity of the solution, (Tarantola 1987). Buland and Omre (2002a) treats the current problem in the Gaussian framework. A stationary Gaussian random field prior, give preference towards smooth solutions. At the geological scales considered in seismic exploration, the earth frequently have slow variations within a layered structure. Hence there appears to be a conflict between a stationary Gaussian random field prior, and the phenomenon under study.

In the current work, a Bayesian methodology is used to solve the problem of seismic inversion. This requires a prior distribution for the material parameters and a likelihood for the observations. To account for the layered structure, the prior distribution is described by a superposition of Gaussian and Cauchy random fields. Both these types of random fields can be constructed by the theory of independently scattered random measures (Rajput and Rosinski 1989). The likelihood model in the current paper is identical to the one used in Buland and Omre (2002a).

The objective of modern Bayesian inference is to explore the posterior distribution, which is formally proportional to the product of the prior and the likelihood. In the current approach this is done by sampling. The samples represent the space of uncertainty with respect to the inversion. In the current presentation the samples are combined to a single estimate using the loss criterion of absolute deviation.

There are alternative Bayesian approaches to modeling of layered structures. Common alternatives are to use a Bernoulli Gauss prior (Mendel 1983), or to model the layering as a point process and mark each point with a random leap size (Malinverno and Leaney 2000). There are also non-Bayesian approaches to inverse problems. Traditional approaches such as quadratic regularization (Tikhonov 1963) and filtering of singular values (Bertero 1989; Hansen 1998) are formally equivalent to Gaussian random field priors (Tarantola 1987; Whaba 1990), and suffer the same deficiencies. However, resent development in computational harmonic analysis, allows for reconstruction by wavelet-vaguelette decomposition (Donoho 1995; Abramovich and Silverman 1998). For a general class of function spaces, the reconstruction adaptively obtain the minimax rate of convergence in the zero noise limit within a log term. This approach is not pursued in the current paper, this is partially because the inverse problem that arise in the current setting does not have sufficient regularity, i.e. the singular values does not have a power law decay.

The data collection procedure and geophysical aspects of the the likelihood model is discussed in Section 2. The statistical construction of the prior and the likelihood is presented in Section 3. The posterior distribution is developed in Section 4, together with the sampling algorithm. In Section 5 the methodology is compared to the standard Gaussian theory in an example where synthetic seismic is generated based on material parameters observed in a well at the Sleipner Øst Field. Section 6 presents inversion of a seismic inline from the Sleipner Øst Field. Lastly, a discussion of the results is included in Section 7.

2 Data collection and geophysical model

The current section gives a brief introduction to practical aspects regarding the data collection and the geophysical assumptions used in the current work. This is all standard methodology in the geophysical community. A more detailed discussion on an introductory level can be found in Sheriff and Geldart (1995).

The material parameters of the earth considered in this work are α being the pressure wave velocity; β being the shear wave velocity; and ρ being the density. These parameters are sufficient to characterize an isotropic elastic medium. Other sets of three parameters are frequently used by geophysicists, but there is an one to one relation between different choices. Below each point at the surface, the material parameters are described by the time profile relative to the reflected wave, $\{\alpha(t), \beta(t), \rho(t)\}$. Figure 1 display the material parameters observed in a well at the Sleipner Øst Field. The observed depth profile is converted to a time profile. The conversion between time and depth is a standard problem in seismic inversion, but is not discussed here.

The seismic data that will be considered in the current presentation is recorded in the North Sea above the Sleipner Øst Field. In a marine seismic survey an

air gun attached to a ship, generates an acoustic wave. The wave propagates through the earth, and is reflected due to contrasts in the sub sea rock formations. The reflected signal propagates back to the surface and is recorded in several hydrophones located in a cable being towed by the ship. The hydrophones record the amplitude of the reflected pressure field as a function of time. A seismic trace is a strain or a pressure amplitude as a function of time. A collection of seismic traces is denoted a gather. The collection of seismic traces recorded in the hydrophones after one shot with the air gun is denoted a common shot point gather, for obvious reasons. In a seismic survey several common shot point gathers are collected, these gathers are further processed. Buland and Omre (2002a) lists 24 different steps of the processing sequence. The processed data are in the form of common depth point gathers, CDP gathers for short. In a CDP gather, the seismic signal correspond to the amplitudes in reflections occurring below one location at the surface, the distance from this location to the shot location is denoted the offset. The processing objective is to make the traces in the CDP gather correspond to primary reflections from a locally vertical 1D-earth model. In a vertical 1D-earth model, the material parameters of the earth is assumed to vary only with depth. The assumption of a vertical 1D-earth model is local, hence material parameters below different locations at the surface may vary at a larger scale.

The seismic signal in CDP gathers used in the inversion, is indexed by angle, θ, in addition to time, t, and are hence denoted angle gathers. The ray path of a wave is defined as the normal vector of the wave front. The angle reference, θ, indicate the angle between the ray path and the vertical line in the point where the reflection occur. Figure 2 show three ray paths that have a common angle of incidence to the vertical line. The amplitudes from these reflections will have the same same angle reference in the angle gather. As the time increases in the angle gather data must be collected at larger offset to keep the angle θ fixed. The time reference in the angle gather defines the depth to the reflecting point in terms of the zero offset reflection time. Figure 3 show three ray paths that have the same depth to the reflection. The amplitudes from these reflections will have the same same time reference in the angle gather. The time reference in the angle gather does not correspond to the physical reflection time. The physical reflection time increase with an increasing offset since the length of the ray path increase, see Figure 3.

A seismic inline of 176 angle gathers, each containing seismic traces for nine angles, were recorded and processed as a part of a seismic survey above the Sleipner Øst Field. Figure 4 shows a time window of the average trace for each angle gather in this inline. The averaging of traces in a gather is denoted stacking in the geophysical terminology. In Figure 5 one of the collected angle gathers is displayed. This gather has approximately the same surface coordinates as the well for which the material parameters are observed, see Figure 1.

The seismic signal in an angle gather can for each angle, θ, be modeled as a

convolution of a wavelet and seismic reflectivity corresponding to this angle,

$$d(\theta,t) = s_\theta * c_{PP}(\theta,t) + e_m(\theta,t)\ , \tag{1}$$

with t being zero offset reflection time; s_θ being a wavelet specific to the angle θ; $c_{PP}(\theta,t)$ being the seismic reflectivity for reflections occurring at the angle θ; and $e_m(\theta,t)$ being model error. For each time coordinate, the seismic reflectivity measure the strength of the reflection from this particular point. The subscript of c_{PP}, indicates that this is the reflectivity of a propagated pressure wave to a reflected pressure wave. There are also reflection coefficients involving shear waves. Only the pure pressure wave reflections are considered here, since only pressure measurements are recorded in the hydrophones.

The Zoeppritz equation describes the dependence between the local material parameters, $\{\alpha(t),\beta(t),\rho(t)\}$, and the seismic reflectivity, $c_{PP}(\theta,t)$, for any angle of the ray path in the 1D-earth model. Stolt and Weglein (1985) introduce a time continuous weak contrast approximation to the Zoeppritz equation. By additional assumptions defined below the dependencies of the angle and the material parameter separates for the seismic reflectivity, yielding the relation,

$$c_{PP}(\theta,t) = a_\alpha(\theta)\frac{d}{dt}\ln\alpha(t) + a_\beta(\theta)\frac{d}{dt}\ln\beta(t) + a_\rho(\theta)\frac{d}{dt}\ln\rho(t) + e_c(\theta,t) \tag{2}$$

with $e_c(\theta,t)$ being model error; and

$$\begin{aligned} a_\alpha(\theta) &= \frac{1}{2}(1+\tan^2\theta)\,, \\ a_\beta(\theta) &= -4\frac{\beta_0^2}{\alpha_0^2}\sin^2\theta\,, \\ a_\rho(\theta) &= \frac{1}{2}\left(1-4\frac{\beta_0^2}{\alpha_0^2}\sin^2\theta\right)\,, \end{aligned} \tag{3}$$

with β_0 and α_0 being constants. The assumption made in Expression (2) is that the ratio of $\beta(t)/\alpha(t)$ can be approximated by β_0/α_0. Buland and Omre (2002a) demonstrate that the modeling error due to this assumption is small compared with the typical noise level in real seismic data. Note that the constant ratio only is assumed in order to approximate Expression (2) and will not carry through to the final estimates. Buland and Omre (2002a) formulate a slightly more general approximation by allowing the ratio to have a preselected time dependence. This type of time dependence could also be included in the current formulation.

The convolutional model and the weak contrast approximation, see Expression (1) and (2), has a limited range of validity. The crucial point being that the seismic preprocessing achieve its goal within a reasonable error. The largest contrast in the material parameters that is observed in the well at the Sleipner Øst Field occur about 2380 ms and have the magnitude 0.3. This is close to

the limit of both the convolutional model and the weak contrast approximation. Below this contrast the angle and time index in the angle gather may be disoriented since the rays bend at the boundary and this effect is not fully accounted for in the preprocessing. Also for large angles surface waves may form in the boundary layer and create additional noise.

The objective is now to reconstruct the material parameters, $\alpha(t)$, $\beta(t)$ and $\rho(t)$ based on the seismic traces in the corresponding angle gather. The material parameters below each location is inverted independently.

3 Statistical model

Bayesian inference requires a statistical model for both the prior and the likelihood. The material parameters are defined on a continuous domain and hence the prior should be modeled by random functions. In addition the prior model should be sufficiently flexible to capture essential characteristics of the material parameters. The likelihood includes the physical link between the observations and the material parameters, and the statistical model for the errors.

3.1 The prior model

The prior model is defined for the logarithm of the material parameters:

$$[\,\alpha_L(t),\,\beta_L(t),\,\rho_L(t)\,] = [\,\ln\alpha(t),\,\ln\beta(t),\,\ln\rho(t)\,]\,.$$

This parameterization guarantees the the material parameters to be positive and is convenient due to the linear relation to the likelihood, see Expression (2). The prior model is described by the locationwise relations;

$$\begin{aligned}
\alpha_L(t) &= \alpha_L^0 + b_\alpha^C \varepsilon_C(t) + b_\alpha^1\,\varepsilon_1(t) + b_\alpha^2\,\varepsilon_2(t) + b_\alpha^3\,\varepsilon_3(t)\,,\\
\beta_L(t) &= \beta_L^0 + b_\beta^C\,\varepsilon_C(t) + b_\beta^1\,\varepsilon_1(t) + b_\beta^2\,\varepsilon_2(t) + b_\beta^3\,\varepsilon_3(t)\,,\\
\rho_L(t) &= \rho_L^0 + b_\rho^C\,\varepsilon_C(t) + b_\rho^1\,\varepsilon_1(t) + b_\rho^2\,\varepsilon_2(t) + b_\rho^3\,\varepsilon_3(t)\,,
\end{aligned} \tag{4}$$

with $\varepsilon_C(t)$ being a centered Cauchy random process; $\varepsilon_1(t), \varepsilon_2(t)$ and $\varepsilon_3(t)$ being centered Gaussian random processes; the b's being scale parameters describing the locationwise dependencies between the material parameters; and α_L^0, β_L^0, and ρ_L^0 being Gaussian random variables centered at the median values for $\alpha_L(t)$, $\beta_L(t)$ and $\rho_L(t)$ respectively. All random components on the right hand side of Expression (4) are assumed to be independent. The Cauchy process $\varepsilon_C(t)$ primarily model the abrupt changes in material parameters, while the Gaussian processes primarily model the smooth variations.

The random processes involved are defined by smoothing of independently scattered random measures (ISRM),

$$\begin{aligned} \varepsilon_C(t) &= \int \phi_C(t-h)\, dC(h) \\ \varepsilon_j(t) &= \int \phi_j(t-h)\, dW_j(h)\ ; \quad j \in \{\, 1,\, 2,\, 3\,\} \end{aligned} \tag{5}$$

with $dC(h)$ being the Cauchy measure; $dW_j(h)$; $j \in \{\, 1,\, 2,\, 3\,\}$ being independent Wiener measures; and ϕ_j; $j \in \{\, C,\, 1,\, 2,\, 3\,\}$ being kernel functions. A short introduction to ISRM is included in Appendix A, a more rigorous presentation is found in Rajput and Rosinski (1989). It is assumed that $\|\phi'_C\|_1 < \infty$ and $\|\phi'_j\|_2 < \infty$; $j \in \{\, 1,\, 2,\, 3\,\}$, with $'$ denoting differentiation with respect to time. This imply that the derivative fields, $\varepsilon'_j(t)$; $j \in \{\, C,\, 1,\, 2,\, 3\,\}$, are stationary. It is further assumed that $\int \varepsilon_j(t)\, dt = 0$ for $j \in \{\, C,\, 1,\, 2,\, 3\,\}$, the integral being over the region under consideration, hence all the variability regarding the global level is represented by α_L^0, β_L^0, ρ_L^0.

The prior distribution as defined above, is stationary for the derivatives of the logarithm of the material parameters, $\alpha'_L(t)$, $\beta'_L(t)$ and $\rho'_L(t)$. By integrating these and center the integrated field, the logarithms of the material parameters will generally not have a stationary distribution, but have higher variability at both ends of the interval. To prevent this effect, information regarding the increase of level in the material parameters could be supplied. For the Sleipner Øst Field this is done by extracting low frequency information from a well and enforce this as an additional constraint to define the prior.

3.2 The likelihood

The profile of the material parameters, is reconstructed independently below each location, using the seismic traces in the corresponding angle gather. The physical link between the observations and the material parameters is defined by combining Expression (1) and (2). As a result the likelihood is defined by,

$$d_{obs}(\theta,t) = a_\alpha(\theta) s_\theta * \alpha'_L(t) + a_\beta(\theta) s_\theta * \beta'_L(t) + a_\rho(\theta) s_\theta * \rho'_L(t) + e(\theta,t)\,, \tag{6}$$

with t being the zero offset reflection time; θ being the angle reference in the angle gather; $d_{obs}(\theta,t)$ being the observed seismic signal; s_θ and the a's being as in Expression (1) and (3); and $e(\theta,t)$ being the error term. The error term represent both observation errors and model errors, and is modeled as a centered Gaussian random field on $[\,0,\, \pi/2\,] \times \boldsymbol{R}$, being independent of the material parameters. The errors are correlated both in time and angles, but to maintain simplicity it is assumed that the dependencies separates in the covariance function,

$$\mathrm{Cov}\{e(\theta_k,t_i), e(\theta_l,t_j)\} = \sigma_\theta(\theta_k)\, \sigma_\theta(\theta_l)\, \nu_\theta(\theta_k,\theta_l)\, \nu_t(t_i,t_j)\,, \tag{7}$$

with $\sigma_\theta(\theta)$ being angle dependent standard deviation; $\nu_\theta(\theta_k, \theta_l)$ being a correlation function describing the dependencies in angular direction; and $\nu_t(t_i, t_j)$ being a correlation function describing the dependencies in time direction. Note that the likelihood only involve the derivative processes, hence it is invariant to changes in α_L^0, β_L^0, and ρ_L^0.

The likelihood is linear with respect to the random measures $dC(h)$ and $dW_j(h)$; $j \in \{ 1, 2, 3 \}$ in Expression (5). Focusing these random measures in Expression (6), one may write,

$$d_{obs}(\theta, t) = \int K_C(\theta, t-h)\, dC(h) + \sum_{j=1}^{3} \int K_j(\theta, t-h)\, dW_j(h) + e(\theta, t) \quad (8)$$

with $d_{obs}(\theta, t)$ being the observed seismic signal; $e(\theta, t)$ being the error; and

$$K_j(\theta, t) = \Big(b_\alpha^j\, a_\alpha(\theta) + b_\beta^j\, a_\beta(\theta) + b_\rho^j\, a_\rho(\theta) \Big)\ \frac{\partial}{\partial t}(s_\theta * \phi_j); \quad j \in \{ C, 1, 2, 3 \},$$

with s_θ being the seismic wavelet, see Expression (1); ϕ_j being kernel functions, see Expression (5); and the a's and the b's being as in Expression (3) and (4).

4 The posterior

The joint posterior distribution of the random measures, $dC(h)$ and $dW_j(h)$; $j \in \{ 1, 2, 3 \}$, see Expression (5) and (8) are explored in this section. The logarithm of the material parameters can then be found by a linear transform of the measures according to Expression (4) and (5). The likelihood is linear, but the Cauchy measure disturbs the traditional Gaussian-linear machinery.

Formally, let C, W and D denote the Cauchy measure, the Wiener measures and the seismic observations respectively. The posterior distribution is explored by splitting it according to the identity:

$$p(w, c|d) = p(c|d)p(w|d, c). \quad (9)$$

The factor $p(c|d)$ is the distribution of a Cauchy measure under a linear constraint with Gaussian errors. The factor $p(w|d, c)$ is the distribution of the Wiener measure under a linear constraint with Gaussian errors. The first factor can be sampled by McMC algorithms, while the second factor can be evaluated analytically.

4.1 Discretization of the problem

In order to implement the Bayesian methodology on a computer, the problem is discretized. This is done by creating discrete equivalents of the relations

above. Having the prior and the likelihood defined on a continuous domain enables control of discretization and assure consistency if the discretization is refined. The random measures are discretized into independent random seeds, see Appendix A. The random processes involved, are then modeled by the discrete equivalent of Expression (5),

$$\begin{aligned} \boldsymbol{\varepsilon}_C &= \boldsymbol{\Phi}_C \boldsymbol{C} \\ \boldsymbol{\varepsilon}_j &= \boldsymbol{\Phi}_j \boldsymbol{W}_j \,; \quad j \in \{\, 1,\, 2,\, 3 \,\} \;, \end{aligned} \tag{10}$$

with $\boldsymbol{\Phi}_j$; $j \in \{\, C,\, 1,\, 2,\, 3 \,\}$ being a matrix representing the convolving kernels, ϕ_j, in Expression (5); $\boldsymbol{C}$ being iid Cauchy seeds; and $\boldsymbol{W}_j$; $j \in \{\, 1,\, 2,\, 3 \,\}$ being three independent sets of iid Gaussian seeds. For later reference let τ and σ denote the scale of Cauchy seeds and Gaussian seeds respectively.

The discrete equivalent of Expression (8), may be written as

$$\boldsymbol{d} = \boldsymbol{K}_{\mathrm{C}} \boldsymbol{C} + \boldsymbol{K}\, \boldsymbol{W} + \boldsymbol{e} \tag{11}$$

with $\boldsymbol{d}$ being the discretized seismic signal in the angle gather; $\boldsymbol{C}$ and $\boldsymbol{W}^T = [\boldsymbol{W}_1^T,\, \boldsymbol{W}_2^T,\, \boldsymbol{W}_3^T\,]$ being the random seeds, see Expression (10); $\boldsymbol{K}_{\mathrm{C}}$ and $\boldsymbol{K}^T = [\,\boldsymbol{K}_1^T,\, \boldsymbol{K}_2^T,\, \boldsymbol{K}_3^T\,]$ being matrices representing the kernels in Expression (8); and $\boldsymbol{e}$ being the discretized error term. The error vector have a multi Gaussian distribution centered at zero and covariance $\boldsymbol{\Sigma}_E$. Exact definitions of the random variables $\boldsymbol{d}$ and $\boldsymbol{e}$ and the matrices $\boldsymbol{K}_{\mathrm{C}}$ and $\boldsymbol{K}$ are given in Appendix B.

The objective is now to sample the posterior distribution of the random seeds $\boldsymbol{C}$ and $\boldsymbol{W}$. Given a sample from this distribution, a sample of the material parameters are obtained by using Expression (4) and (10).

4.2 The posterior Cauchy seed

When the focus is on the Cauchy seed, Expression (11) can be restated as,

$$\boldsymbol{d} = \boldsymbol{K}_{\mathrm{C}} \boldsymbol{C} + \boldsymbol{e}_N$$

with $\boldsymbol{d}$, $\boldsymbol{K}_{\mathrm{C}}$, $\boldsymbol{C}$ being as in Expression (11); and $\boldsymbol{e}_N = \boldsymbol{K}\, \boldsymbol{W} + \boldsymbol{e}$, hence the error structure is altered,

$$p(\boldsymbol{e}_N) = N_{mn}\,(\boldsymbol{0}, \boldsymbol{\Sigma}_N)\,,$$

with $\boldsymbol{\Sigma}_N = \sigma^2 \boldsymbol{K}\boldsymbol{K}^T + \boldsymbol{\Sigma}_E$ being the covariance matrix. The posterior can hence be written as:

$$\begin{aligned} p(\boldsymbol{c}|\boldsymbol{d}) = \quad & const \times \left[\prod_{i=1}^{n} \left(\tfrac{1}{\pi\tau(1+(c_i/\tau)^2)}\right)\right] \\ & \times \exp\left\{-\tfrac{1}{2}(\boldsymbol{K}_{\mathrm{C}}\boldsymbol{c} - \boldsymbol{d})^T \boldsymbol{\Sigma}_N^{-1} (\boldsymbol{K}_{\mathrm{C}}\boldsymbol{c} - \boldsymbol{d})\right\} \;. \end{aligned} \tag{12}$$

The posterior distribution of $\boldsymbol{C}$ is multi modal. By differentiation of Expression (12) it is easily found that any local mode $\boldsymbol{c}^{\text{LM}}$ satisfy the relation,

$$\boldsymbol{c}^{\text{LM}} = \left(\boldsymbol{K}_{\text{C}}^T \boldsymbol{\Sigma}_N^{-1} \boldsymbol{K}_{\text{C}} + \boldsymbol{D}(\boldsymbol{c}^{\text{LM}}) \right)^{-1} \boldsymbol{K}_{\text{C}}^T \boldsymbol{\Sigma}_N^{-1} \boldsymbol{d} \,, \tag{13}$$

with $\boldsymbol{D}(\boldsymbol{c}^{\text{LM}})$ being a diagonal matrix with $\boldsymbol{D}(\boldsymbol{c}^{\text{LM}})_{ii} = \frac{2}{\tau^2 + (c_i^{\text{LM}})^2}$. The expression for $\boldsymbol{c}^{\text{LM}}$ above should be compared with the corresponding for a Gaussian prior, for which $\boldsymbol{D}(\boldsymbol{c}^{\text{LM}})_{ii} = \frac{1}{\tau_G^2}$. The maximum posterior estimator correspond to linear shrinkage in the Gaussian model, whereas in the current model the maximum posterior correspond to nonlinear shrinkage.

The modes of the posterior distribution, $p(\boldsymbol{c}|\boldsymbol{d})$, are located close to subspaces constructed by setting most of the coefficients of $\boldsymbol{C}$ to zero. This is due to the sparse structure of the Cauchy seed, as is pointed out in Appendix A. This fact allows for efficient sampling based on the multi directional Gibbs sampler (Liu and Sabatti 2000). Details of the algorithm is in Appendix C and E.

4.3 The posterior Gaussian seed

When the value of the Cauchy seed is given, $\boldsymbol{C} = \boldsymbol{c}$, Expression (11) can be expressed as,

$$\boldsymbol{d} - \boldsymbol{K}_{\text{C}} \boldsymbol{c} = \boldsymbol{K}\boldsymbol{W} + \boldsymbol{e} \,.$$

The conditional posterior can be calculated explicitly ,

$$p(\boldsymbol{w}|\boldsymbol{d}, \boldsymbol{c}) = N_{3n} \left(\boldsymbol{\mu}_{\text{w}}, \boldsymbol{\Sigma}_{\text{w}} \right),$$

with

$$\boldsymbol{\mu}_{\text{w}} = \sigma^2 \boldsymbol{K}^T \left(\sigma^2 \boldsymbol{K}\boldsymbol{K}^T + \boldsymbol{\Sigma}_E \right)^{-1} (\boldsymbol{d} - \boldsymbol{K}_{\text{C}} \boldsymbol{c})$$

$$\boldsymbol{\Sigma}_{\text{w}} = \sigma^2 \boldsymbol{I} - \sigma^4 \boldsymbol{K}^T \left(\sigma^2 \boldsymbol{K}\boldsymbol{K}^T + \boldsymbol{\Sigma}_E \right)^{-1} \boldsymbol{K}$$

Note in particular that the posterior covariance does not depend on the value of the Cauchy seed, hence if the maximum posterior solution for the seeds is sought, the search may be done sequentially in Expression (9).

Since the full conditional distribution is Gaussian, there are several standard ways to sample the posterior distribution in this case. The approach used here is direct simulations, details of the algorithm is in Appendix E.

Because the Gaussian seeds are easy to sample when the Cauchy seed is given, several Gaussian seeds are generated for each Cauchy seed.

5 Comparison to a pure Gaussian model

The model defined in the current work is compared with a model having a purely Gaussian prior. The test case is based on material parameters observed in the well at the Sleipner Øst Field, see Figure 1, synthetic seismic observations are generated based on this profile. In Figure 1 the depth profile is converted to a time profile. The region under consideration is the time interval from $2000 - 2400$ ms. The drilling stopped after 2390 ms below this depth the value of the material parameters is fixed at a constant level.

The prior parameter values are estimated from the well observations. The estimation procedure is described in the next section and Appendix D. The level and scale parameters, see Expression (4), are listed in Table 1 and 2, the derivative of the kernels, ϕ_j'; $j \in \{C, 1, 2, 3\}$, see Expression (5), are displayed in Figure 6 and 7 for the Cauchy model and the pure Gaussian model respectively. A random effect could be added to the level parameters, but this serves no purpose in this example since the seismic observations are unaffected by a change in these parameters.

The likelihood is modeled by the linearized 1D-earth model, see Expression (1), (2) and (3), using $\beta_0/\alpha_0 = 0.5$. The data is collected for nine equispaced angles from 5° to 37°. The smoothing wavelet of Expression (1) is assumed to have the functional form

$$s_\theta(t) = a(\nu)\,(1.5 - 2\pi^2\nu^2t^2)\exp\{-\pi^2\nu^2t^2\}, \tag{14}$$

with $a(\nu)$ being an amplitude scale; and $\nu = 25$ Hz being the peak frequency. The wavelet is independent of angle and the amplitude is selected such that $\|s_\theta\|_2 = 1$. This is a Ricker like wavelet, see Sheriff and Geldart (1995), modified such that the low frequency components are larger. This is to avoid large fluctuations in the low frequent components in the solution, and is done instead of redefining the prior to contain lowfrequent information.

The error structure have a white and a colored component. The white noise component contribute 1% of the variance, and the colored component contribute the remaining 99%. The colored component has the form given in Expression (7). The random error in time direction is obtained by convolving the seismic wavelet, see Expression (14), with the Wiener measure, this defines $\nu_t(t_i, t_j)$. The error correlation in the direction of angles is given by,

$$\nu_\theta(\theta_k, \theta_l) = \exp\{-3|\theta_k - \theta_l|/\eta_\theta\} \tag{15}$$

with $\eta_\theta = 60°$ being the length scale for the correlation. The standard deviation, $\sigma_\theta(\theta)$, is assumed constant as a function of angle and is 0.002. A discussion of the correlation structure is included in the next section.

The synthetic observations are obtained by applying the linearized forward model to the parameters from the well logs, and add errors according to the

likelihood. The synthetic observations are displayed in Figure 8. The signal to noise ratio, measured as the ratio of the squared L^2 norms, is about 64. Under the loss criterion of absolute deviation the estimate is the median of the marginal posterior distribution for the relevant parameter.

The estimates in the Cauchy model are evaluated using 800 sampled Cauchy seeds, sampling five Gaussian seeds for each Cauchy seed. The results from the Gaussian model can be obtained analytically. The estimates and the true parameter values are displayed in Figure 9 and 10 for the Cauchy model and the pure Gaussian model respectively. Figure 11 and 12 show the error in the estimates together with a 90% credibility interval.

Evaluating the estimates by eye, both methods keep the true value reasonable well within the error margin and captures the main features in the velocities, $\alpha(t)$ and $\beta(t)$, whereas neither model resolve the density, $\rho(t)$, satisfactory. The biggest visual difference appears at the leap in the parameter values at 2380 ms. For the Cauchy model the leap is present in the estimate, whereas in the Gaussian model it is smoothed over a region, compare Figure 9 and 10. The error for the Cauchy model seems unaffected by the leap, but the smoothing in the pure Gaussian model results in large fluctuations in the error around 2380 ms, compare Figure 11 and 12. A similar effect is observed around the peak at 2310 ms.

When the estimates are evaluated by the criterion of absolute deviation the Cauchy model produces a better fit for the velocity estimates but there is no gain for the density. For the material parameters in Figure 1 the risk is estimated by averaging the loss for 100 errors simulated according to the likelihood. The overall risk improvement is 7% for the velocity estimates. No improvement is noted for the rock density. If only the interval from 2000 ms to 2250 ms is considered there is essentially no difference between the two models. The main advantage of the Cauchy model is observed in the region 2250 ms to 2400 ms, where the velocities have large fluctuations. In this region the estimates are improved by 14 % and 10 % for $\alpha(t)$ and $\beta(t)$ respectively.

The Cauchy model reduces the average error margin in the 90% credibility interval by 7% for the velocities $\alpha(t)$ and $\beta(t)$. However in the regions with large fluctuations, the error margins for the Cauchy model is larger than for the pure Gaussian model. Note in particular the region around 2310 ms where the credibility interval increases in the Cauchy model. Note also the characteristic peaks in the credibility intervals in the regions where the leaps are, i.e. around 2310 ms and 2380 ms. This indicate uncertainty in the jump location. In the top region between 2000 ms to 2250 ms the error margin in the 90% credibility interval is about 10% shorter in the Cauchy model compared with the pure Gaussian model.

6 Sleipner Øst Field

In this section the seismic inline of 176 angle gathers collected above the Sleipner Øst Field, see Figure 4, is inverted independently in each location. The inversion of the Sleipner data is based on the procedure defined above, using the observed part of the well log in Figure 1 to fit parameters in the prior model, and the colocated angle gather, see Figure 5, to estimate the parameters in the likelihood. Normal plots of the derivative of the logarithmic material parameters from the well log are displayed in Figure 13. The figure clearly illustrate the heavy tailed nature of the phenomenon, hence the need for a non-Gaussian model in this case study.

By assuming the processes defined as the derivative of the logarithmic material parameters to be ergodic, averages under the marginal distribution can be evaluated by time averages of these processes. The scale parameters in Expression (4) can hence be estimated by a modified method of moments, keeping the scale of the Cauchy field at fixed ratios. The final estimates are listed in Table 1. Details about the estimation procedure are left to Appendix D. The Cauchy process is primarily modeling the layered structure, hence the smoothing kernel for the derivative process is fixed as a Dirac at the level of grid resolution. The kernel is displayed in Figure 6 (a). The Gaussian processes, $\varepsilon_1'(t)$, $\varepsilon_2'(t)$ and $\varepsilon_3'(t)$, are then estimated by subtracting an estimate of the Cauchy field, and decouple the processes by the inverse of the estimated scale matrix for the Gaussian processes. The correlation function for each of the Gaussian processes are then fit by eye to an empirical estimate. A more advanced technique combining the methodology of Appendix D with tapering (Fodor and Stark 2000) could be developed, but is not believed to give any significant advantage in the current application. The estimated kernels, being the symmetric square root of the correlation operator, are displayed in Figure 6 (b)-(d). The kernels are those used in the example of the previous section.

In the likelihood the seismic wavelet and the noise covariance must be estimated. This is an inverse problem by itself. Buland and Omre (2002b) uses a Bayesian approach with vague priors to estimate the parameters. A slightly modified method is used here. Before the estimation is performed the well log and the seismic traces are aligned in time such that the highest peak of the seismic traces match the leap in parameter values at 2380 ms. The estimated wavelets are displayed in Figure 14.

The residuals after the wavelet estimation have two components of error, being due to observation errors in the well logs and in the seismic data. Only the sum of the errors is identifiable. Hence a subjective choice must be made regarding the seismic error structure. The variance of the error is modeled by a colored component of 99 % and a white component of 1 %. The colored component of the error is mainly due to effects related to the approximate model and imperfections of the seismic processing, hence the error is on the same scales as

the data. The time correlation of the error is hence obtained by convolving an average wavelet with white noise, this defines $\nu_t(t_i, t_j)$. The angle correlation of the error is chosen to be a first order exponential correlation function, see Expression (15), with length scale $\eta_\theta = 30°$. The standard deviation of the error $\sigma_\theta(\theta)$ is chosen to be constant and have the value 0.35. This correspond to a signal to noise ratio ranging from five to two since the energy in the wavelets vary with angle.

The seismic wavelet contain only intermediate frequencies. This gives large fluctuations in the sampled values due to uncertainty in the low frequencies. In Gaussian models it is common to extract the low frequency components from the well and center the samples and estimates around this mean value. This approach is not optimal for the Cauchy model since it interfere with the structure of the prior distribution. In the current approach the low frequency content of the well is included by extracting the information below 10 Hz from the well log. The prior distribution is now redefined. The new prior distribution is the distribution previously defined conditioned to observations of the low frequency components according to the well observations. Figure 15 show the pointwise median and 90% credibility interval for the material parameters in the well location when the low frequency information is included.

The estimates are based on 200 samples of the Cauchy seeds, sampling 10 Gaussian seeds for each Cauchy seed. The mixing of the sampling algorithm is briefly commented in Appendix C and appears to be satisfactory. A larger number of samples would have been desirable to reduce the Monte Carlo variation in the estimates. The parameters are reconstructed in the region between 2050 ms and 2450 ms to avoid boundary effects due to missing observations.

Figure 16 displays the estimates below the well location together with 90% credibility interval. Comparing this figure with Figure 15 it is seen that the seismic only carry a moderate amount of information regarding the parameters, however the estimates contain more details than the prior and the credibility intervals are generally shorter than in the prior. In Figure 17 the current estimates are compared with the well logs. The estimates have a reasonable good correspondence, except from the peak at about 2310 ms that bears no effect in the estimate of $\alpha(t)$. Contrary to the synthetic example, the estimate is smoothed in the region containing the leap. This is due to uncertainty in the jump location. Most individual samples have one jump at about the right position, but this location varies slightly between samples, this reflects the multi modality of the posterior distribution. When the posterior distribution is multi modal, such as in the current case, it is in general impossible to find one estimate that both have a characteristic shape and represent average properties of the posterior. The global mode of the posterior distribution is an alternative estimator. This estimator will reproduce a leap in the parameter value, and hence be visually attractive. It would however have a worse performance if measured by average quantities.

Figure 18 show the the 90% credibility interval for the error and the actual error for the well. The smoothing effect of the estimate around the leap produces a large error in this location, but even this large error is within the credibility interval due to the characteristic peak in the credibility interval around the leap. The peak in the credibility interval indicate the uncertainty in jump location.

The final estimates of α, β and ρ for all the 176 gathers are displayed in Figure 19(a)-(c). The leap value between 2350 ms and 2450 ms dominates the picture, but other details are also present. The Monte Carlo variation of the estimates causes some disturbance to the pictures in particular for the density estimate, see Figure 19(c). For the estimate of β, see Figure 19(b), the the estimate is threshold at 3500 m/s in order to represent the contrasts in the estimate better, the estimate exceed this value in some areas below the leap these areas are colored red. the maximum value of β is about 4250 m/s and occur in the read area at the leap boundary for gather 43. Below the leap especially in the lower left corner of the figure the estimates fluctuate. The basis of these fluctuations are present i the data, but may be due to imperfections of the preprocessing. There are several sources of errors. If time axis is shrunk to much in the preprocessing energy is migrated to higher frequencies, if the effect of geometrical spreading and absorption is over compensated the signal is amplified, also the velocity dependence of the references in the angle gather may be problematic. In general the estimates are reasonable.

7 Discussion with conclusions

The prior model is defined in a consistent way by the use of independent scattered random measures. A superposition of Cauchy and Gaussian processes models a layered structure with slow variations within each layer. Compared to the more common Bernoulli Gauss model, the Cauchy model introduces the layering without introducing dichotomy explicitly, and is defined independent of grid. The current model is an alternative to modeling the layers by point processes. The likelihood model is based on well founded geophysical principles.

The estimator is evaluated by stochastic simulation. The loss criterion of absolute deviation account for all the generated samples in a robust way. In an example the Cauchy model is compared to a pure Gaussian model. The test example is based on real material parameters observed in the Sleipner Øst Field and synthetic seismic data is generated. For the velocities the Cauchy model is found to reduce the over all risk by 7%. In regions where the material parameters have large fluctuations the estimates of the velocities improve by 14 % and 10 %. Over all the error margins in the 90% credibility interval is reduced by 7%. In regions where the material parameters are slowly varying the 90% credibility intervals for the velocities are reduced by 10%. The model only have a minor impact on estimation of the rock density. In general the uncertainty of the estimates are well represented in the Cauchy model in particular the

uncertainty in leap locations.

For the Sleipner Øst Field, the model define a prior distribution that account for the layered structure. To restore the material parameters, with reasonable error margins, low frequency information must be included from the well log. In the current approach this is done by redefining the prior distribution to be the conditional distribution when given low frequency information from the well log. The inversion results are satisfactory and the uncertainty is well represented. The uncertainty is however large due to large errors in the data.

Ideally a model that accounts for lateral dependencies should be developed. In such a model the well information could be included in a consistent way. The current model does not immediately generalize to include lateral dependencies. It is however possible to define a spatial model by superposition of stationary Cauchy fields in higher dimensions, such a model would however substantially increase the computational cost and raise additional questions regarding estimation of prior parameters and sampling algorithm.

The Cauchy model might be considered to have too heavy tails, hence unrealistically large leaps in the parameter values may occur according to the prior distribution. The random fields defined above are related to random fields of type $\mathcal{G}$ (Barndorff-Nielsen and Prez-Abreu 2002). Random fields of type $\mathcal{G}$ are very general and offer a broad specter of prior distributions that can be investigated. In particular multivariate normal inverse Gaussian distributions is a class of flexible distributions that bridges the gap between multivariate Cauchy distributions and Gaussian distributions. To utilize such random fields as priors is a topic for further research.

Acknowledgments

This work was supported by the Research Council of Norway. The author thanks Henning Omre, Arild Buland and Håkon Tjelmeland for helpful comments, and Statoil and the Sleipner licence (Statoil, Exxon/Mobil, Norsk Hydro and TotalFinaElf) for permission to publish this paper

References

Abramovich F. and Silverman B.W. (1998), "Wavelet decomposition approaches to statistical inverse problems," *Biometrika*, 85, 115-129.

Barndorff-Nielsen, O.E. and Prez-Abreu, V. (2002) "Multivariate Type $\mathcal{G}$ Distributions " To appear in Theory Prob. Its Appl

Bertero, M. (1989), "Linear inverse and ill-posed problems" Advances in electronics and electron physics, Vol. 75, pp 1-120.

Buland, A. and Omre, H. (2002a), "Bayesian linearized AVO inversion" To appear in Geophysics.

Buland, A. and Omre, H. (2002b), "Bayesian wavelet estimation from seismic and well data " To appear in Geophysics.

Donoho, D.L. (1995), "Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition," *Applied and computational harmonic analysis*, 2, 101-126.

Fodor, I.K., Stark, P.B. (2000) "Multitaper Spectrum Estimation for time series with gaps" IEEE Transactions on signal processing Vol. 48, 3472-3483.

Hansen, P.C. (1998) "Rank-deficient and discrete ill-posed problems," SIAM Monographs on Mathematical Modeling and Computation. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA

Liu, J.S. and Sabatti, C. (2000) "Generalized Gibbs Sampler and Multigrid Monte Carlo for Bayesian Computation", *Biometrika*, 87, 353-369.

Malinverno, A. and Leaney, S. (2000) A Monte Carlo method to quantify uncertainty in the inversion of Zero-Offset VPS data.

Mendel, J.M. (1983) "Optimal Seismic Deconvolution: An Estimation-Based Approach", Academic Press, New York.

Rajput, B.S. and Rosinski, J. (1989) "Spectral Representation of Infinitely Divisible Processes." Probability Theory and related fields. Vol 82, 451-487.

Sheriff, R.E. and Geldart,L.P. (1995) "Exploration Seismology" , 2 ed., Cambridge University Press.

Stolt, R. H. and Weglein, A.B., (1985), "Migration and inversion of seismic data"; Geophysics Vol. 50, pp 2458-2472.

Tarantola, A. (1987), "Inverse problem Theory"; Elsevier.

Tikhonov, A. (1963), "Solution to incorrectly formulated problems and the regularization method," *Soviet Math. Doklady*, 5, 1035-1038.

Vanmarcke, E. (1983), "Random fields"; The MIT press.

Wahba, G. (1990), "Spline models for Observational data", *NSF-CBMS Regional Conference in Mathematics*, (Vol. 59); Philadelphia; SIAM.

A Independently scattered random measures

Independently scattered random measures offers a way of constructing prior distributions for functions defined on a continuous domain in $\mathbf{R}^d$. The presentation below is motivated from the point of application, a more rigorous presentation is found in Rajput and Rosinski (1989).

Definition 1 (Independently scattered random measures,(ISRM)) *An independently scattered random measure, $Z \in \mathbf{R}^d$ is a set of random variables indexed by the Borel sets in $\mathbf{R}^d$, such that for any sequence of disjoint Borel sets $\{\mathcal{A}_i\}_{i=1}^{\infty}$ in $\mathbf{R}^d$ the two properties hold.*

$$\begin{aligned} i) \quad & Z(\mathcal{A}_i)\ ;\ i = 1, 2, \ldots\ ;\ \textit{are independent,} \\ ii) \quad & Z(\textstyle\bigcup_i \mathcal{A}_i) = \sum_i Z(\mathcal{A}_i) \qquad \textit{a.s.,} \end{aligned}$$

with a.s. denoting almost sure convergence of the series.

The Wiener measure and the Cauchy measure will be defined next. Let $\stackrel{\mathcal{D}}{=}$ denote equality in distribution, and $|\cdot|$ denote the Lebesgue measure on $\boldsymbol{R}^d$.

The Wiener measure, can be defined by $i)$ and $ii)$ in addition to

$$iii)_w \quad Z(\mathcal{A}_i) \stackrel{\mathcal{D}}{=} \text{Gauss}(0, |\mathcal{A}_i|),$$

with $\text{Gauss}(0, \sigma^2)$ being a Gaussian distributed random variable with zero mean and variance σ^2. The Wiener measure is a signed random measure on $\mathbf{R}^d$. In what follows W is used to refer to this measure.

The Cauchy measure, can be defined by $i)$ and $ii)$ in addition to

$$iii)_c \quad Z(\mathcal{A}_i) \stackrel{\mathcal{D}}{=} \text{Cauchy}(|\mathcal{A}_i|),$$

with $\text{Cauchy}(\tau)$ being a centered Cauchy distributed random variable with the scale factor τ. The density of a $\text{Cauchy}(\tau)$ random variable is:

$$p_C(x;\tau) = \frac{1}{\pi\tau\left(1 + \left(\frac{x}{\tau}\right)^2\right)} \quad , \qquad x \in \boldsymbol{R}, \tau > 0\ .$$

The Cauchy measure is a signed random measure on $\mathbf{R}^d$. In what follows C is used to refer to this measure.

A.1 Stationary random fields defined by ISRM

The stationary fields obtained by convolving a kernel ϕ with the ISRM,

$$\varepsilon(t) = \int_{\mathbf{R}^d} \phi(t-h)\, dZ(h), \tag{16}$$

is of particular interest in the current application.

The random fields generated by the Wiener measure and the Cauchy measure will be used for constructing the prior distribution. Let $\|\phi\|_1$ and $\|\phi\|_2$ denote the L^1 and L^2-norms respectively and assume $\|\phi\|_1 + \|\phi\|_2 < \infty$.

A stationary Gaussian random field is defined by

$$\varepsilon_G(t) = \int_{\mathbf{R}^d} \phi(t-h)\, dW(h).$$

According to the standard theory $\varepsilon_G(t)$ have a covariance function uniquely defined by ϕ (Vanmarcke 1983).

A stationary Cauchy random field is defined by

$$\varepsilon_C(t) = \int_{\mathbf{R}^d} \phi(t-h)\, dC(h).$$

The term Cauchy field is twofold deserved. Firstly, it is constructed based on the Cauchy measure, secondly all marginal distributions are Cauchy distributed,

$$\varepsilon_C(t) \stackrel{\mathcal{D}}{=} \text{Cauchy}(\|\phi\|_1).$$

Linear transforms of stationary fields defined by Expression (16), are given by transforming the kernel ϕ correspondingly. Let K denote the linear transform and apply this transform to the random field, then

$$K\varepsilon(t) = \int_{\mathbf{R}^d} K\phi(t-h)\, dZ(h),$$

with $K\varepsilon$ being the transformed field; and $K\phi$ being the linear transform applied to the kernel ϕ. Appropriate regularity conditions must apply to ϕ and K in order to make $K\varepsilon$ well defined. In the current application the linear transforms of interest are integration, convolution, and differentiation.

A.2 Discretization of ISRM

In the current article independent scattered random measures are used to define prior distributions for functions on a continuous domain. Having a continuously defined prior, enables control of the discretization error, and guarantees stability of the discretization as the resolution increases. An independently scattered random measure is discretized into independent random seeds by integrating over small volumes. For $t \in \mathbf{R}^d$ use the multi index notation to denote $t_\alpha = \alpha \cdot \Delta t = (\alpha_1 \Delta t_1, \alpha_2 \Delta t_2, ..., \alpha_d \Delta t_d)$ for $\alpha \in \boldsymbol{Z}^d$, and let $|\Delta t|$ denote the Lebesgue measure of the volume element Δt.

Discretization of the Wiener measure results in,

$$W_\alpha = \int_{t_\alpha}^{t_\alpha + \Delta t} dW(h),$$

with W_α being an iid sequence of random variables such that

$$W_\alpha \stackrel{\mathcal{D}}{=} \text{Gauss}(0, |\Delta t|).$$

For the Wiener measure the standard normal seed is scaled by $\sqrt{|\Delta t|}$.

Discretization of the Cauchy measure results in,

$$C_\alpha = \int_{t_\alpha}^{t_\alpha + \Delta t} dC(h),$$

with C_α being an iid sequence of random variables such that

$$C_\alpha \stackrel{\mathcal{D}}{=} \text{Cauchy}(|\Delta t|).$$

For the Cauchy measure the standard Cauchy seed is scaled by $|\Delta t|$.

In Figure 20 the Wiener measure and the Cauchy measure are visualized by discretized realizations. The figure clearly reveal the underlying structure of the random measures. The Wiener measure distributes the energy equally along the region, while the Cauchy measure concentrates most of the energy in a few locations. The focusing of energy is one of the properties that motivated the use of Cauchy fields in the current application.

B Details regarding discretization of the problem

The observations are collected for the angles θ_k; $k = 1, ..., m$; for each angle the functions are discretized into vectors of length n. For a given angle θ_k Expression (6) then translates into

$$\boldsymbol{d}_k = a_\alpha(\theta_k)\boldsymbol{S}_k\boldsymbol{D}\boldsymbol{\alpha}_L + a_\beta(\theta_k)\boldsymbol{S}_k\boldsymbol{D}\boldsymbol{\beta}_L + a_\rho(\theta_k)\boldsymbol{S}_k\boldsymbol{D}\boldsymbol{\rho}_L + \boldsymbol{e}_k, \qquad (17)$$

with $\boldsymbol{d}_k$ being the discretized seismic traces; the a's being as for Expression (3); $\boldsymbol{S}_k$ being a matrix representing convolution with the wavelet s_{θ_k}; $\boldsymbol{D}$ being a matrix representing differentiation; $\boldsymbol{\alpha}_L$, $\boldsymbol{\beta}_L$ and $\boldsymbol{\rho}_L$ being discretization of the material parameters; and $\boldsymbol{e}_k$ being the discrete error at angle θ_k. The full error vector $\boldsymbol{e}^T = [\boldsymbol{e}_1^T, ..., \boldsymbol{e}_m^T]$ have a multivariate Gaussian distribution,

$$p(\boldsymbol{e}) = N_{nm}(\boldsymbol{0}, \boldsymbol{\Sigma}_E)$$

with $\boldsymbol{\Sigma}_E$ being the covariance matrix. Due to the separability, see Expression (7), the covariance matrix have the form $\boldsymbol{\Sigma}_E = \boldsymbol{\Sigma}_\theta \otimes \boldsymbol{\Sigma}_t$, with $\boldsymbol{\Sigma}_\theta$ and $\boldsymbol{\Sigma}_t$ being $m \times m$ and $n \times n$ matrices describing the error covariance and correlation in direction of angles and time respectively; and $\otimes$ being the Kronecker tensor product.

Focusing on the random seeds $\boldsymbol{C}$ and $\boldsymbol{W}_j$; $j \in \{1, 2, 3\}$ in Expression (10), the likelihood is linear,

$$\boldsymbol{d} = \boldsymbol{K}_{\mathrm{C}}\boldsymbol{C} + \boldsymbol{K}\,\boldsymbol{W} + \boldsymbol{e} \tag{18}$$

with $\boldsymbol{d}^T = [\boldsymbol{d}_1^T, ..., \boldsymbol{d}_m^T]$ being the discretized seismic traces; $\boldsymbol{C}$ and $\boldsymbol{W}^T = [\boldsymbol{W}_1^T, \boldsymbol{W}_2^T, \boldsymbol{W}_3^T]$ being the random seeds; $\boldsymbol{K}_{\mathrm{C}}$ being a $nm \times n$ matrix; $\boldsymbol{K}$ being a $mn \times 3n$ matrix; and $\boldsymbol{e}$ being the error term as defined above. $\boldsymbol{K}^T = [\boldsymbol{K}_1^T, \boldsymbol{K}_2^T, \boldsymbol{K}_3^T]$. The matrices $\boldsymbol{K}_j$; $j \in \{C, 1, 2, 3\}$ all have the same form, the relation is

$$\boldsymbol{K}_j = \begin{bmatrix} a_1^j \boldsymbol{S}_1 \boldsymbol{D} \boldsymbol{\Phi}_j \\ \vdots \\ a_m^j \boldsymbol{S}_m \boldsymbol{D} \boldsymbol{\Phi}_j \end{bmatrix},$$

with $\boldsymbol{S}_k$ being the matrix representing convolution with the wavelet at at angle θ_k;$\boldsymbol{D}$ being a matrix representing differentiation; $\boldsymbol{\Phi}_j$ being matrices representing the kernel functions, see Expression (10); and

$$a_k^j = a_\alpha(\theta_k) b_\alpha^j + a_\beta(\theta_k) b_\beta^j + a_\rho(\theta_k) b_\rho^j\,, \quad j \in \{C, 1, 2, 3\}\,,$$

with the a's on the right hand side being as in Expression (3); and the b's being as in Expression (4).

C The multi directional Gibbs sampler

The multi directional Gibbs sampler is a particular case of the generalized Gibbs sampler (Liu and Sabatti 2000). The generalized Gibbs sampler is a Markov chain based method for sampling a distribution. A Markov chain having the target distribution as stationary distribution is constructed by defining the updating rules.

Let $\boldsymbol{c}^{\mathrm{old}}$ be the current state of the chain; $\{\boldsymbol{v}^l\}_{l=1}^L$ be an over complete pool of basis vectors; and s be a scalar. For the purpose of describing the algorithm, let $p(\cdot)$ be the target density. The multi directional Gibbs sampler is described by the updating rule:

1. Draw l^* uniformly from $\{1, ..., L\}$
2. Draw s^* from $q(s) \propto p(\boldsymbol{c}^{\mathrm{old}} + s \cdot \boldsymbol{v}^{l^*})$
3. Let $\boldsymbol{c}^{\mathrm{new}} = \boldsymbol{c}^{\mathrm{old}} + s^* \cdot \boldsymbol{v}^{l^*}$

In the current application of the multi directional Gibbs sampler, the unit vectors are chosen as translations of the vectors in Figure 21. The density values

of $q(s)$ is known to proportionality by Expression (12),

$$q(s) \propto \left[\prod_{\{i\,:\,v_i^{l^*} \neq 0\}} \frac{1}{\pi\tau \left(1 + \left(\frac{c_i^{\text{old}} + s\, v_i^{l^*}}{\tau}\right)^2\right)} \right] \exp\{-\tfrac{1}{2\sigma_{l^*}^2}[s - \boldsymbol{\mu}_{l^*}^T(\boldsymbol{d} - \boldsymbol{K}_{\text{C}}\boldsymbol{c}^{\text{old}})]^2\},$$

with

$$\sigma_{l^*}^2 = \left[(\boldsymbol{v}^{l^*})^T \boldsymbol{K}_{\text{C}}^T \boldsymbol{\Sigma}_N^{-1} \boldsymbol{K}_{\text{C}} \boldsymbol{v}^{l^*}\right]^{-1},$$

and

$$\boldsymbol{\mu}_{l^*} = \sigma_{l^*}^2 \boldsymbol{\Sigma}_N^{-1} \boldsymbol{K}_{\text{C}} \boldsymbol{v}^{l^*}.$$

The likelihood give essential bounds for the posterior range. The univariate density $q(s)$ is calculated on a dense grid centered at $\boldsymbol{\mu}_{l^*}^T(\boldsymbol{d} - \boldsymbol{K}_{\text{C}}\boldsymbol{c}^{\text{old}})$ and stretching $5\sigma_{l^*}$ to each side. Note that σ_l and $\boldsymbol{\mu}_l$ are specific to each unit vector $\boldsymbol{v}^l$, and that their expressions only depend on the angles for which the traces are observed. Since the same set of angles are used for all gathers, σ_{l^*} and $\boldsymbol{\mu}_{l^*}$ only need to be calculated once for the total study.

In the current work the simulation is initiated in a local mode. The local mode is found by iterating Expression (13). After a burn in of 20 random scans through the pool of basis vectors, a sample is extracted at the end of every 2end random scan through the pool of basis vectors. The mixing for the Cauchy seed in the Sleipner Øst Field is displayed in Figure 22. The Cauchy seed is visualized in the figure. The middle plot show the sampled seed value that correspond to the leap value at 2380 ms. The seeds form the two nearest neighbors at both sides are in the plots above and below, respectively. In the figure three samples are extracted in each random scan through the pool of basis vectors. The figure clearly show how the sampling algorithm move between different modes corresponding to different locations for the leap in the parameter values.

D Estimation of prior scale parameters

In this appendix a methodology for estimating the scale parameters of Expression (4) is supplied. The scaling of the Gaussian processes is the square root of the pointwise covariance matrix for the logarithm of the material parameters given the Cauchy process. This part requires six parameters to be estimated. Let $\boldsymbol{B}_G$ denote a symmetric matrix with the values of the scaling of the Gaussian field. For the Cauchy process three parameters must be estimated.

If all nine parameters are to be determined, the estimates are unstable. This is seen by the fact that if all random processes are Gaussian, only six parameters can be identified. To further reduce the number of parameters, the ratios of the parameters for the Cauchy process are held fixed; $[\,b_\alpha^C,\, b_\beta^C,\, b_\rho^C\,] = \gamma \boldsymbol{v}^T$ with

$\boldsymbol{v}^T = [\,1.0,\, 1.0,\, 0.15\,]$, only the global scale γ is estimated. According to geophysical literature it is reasonable to assume that the first two components are of approximately of the same size, the third component is selected to be about size of $(\rho_L(2500) - \rho_L(2000))/(\alpha_L(2500) - \alpha_L(2000))$. The scale parameters are estimated by a modified method of moments, by computing the sample averages and tune the scale parameters so that the population averages match the sample averages for a given set of functions.

For a stationary random process, $\{\varepsilon(t),\ t \in \boldsymbol{R}\}$, define $\varepsilon_i = \int_{t_i}^{t_i+\Delta t} \varepsilon(s)ds$. The random variables ε_i are then identically distributed. Let ε denote a generic random variable being distributed according to the law of ε_i . If $\{\varepsilon(t),\ t \in \boldsymbol{R}\}$, is ergodic, the sample averages of ε_i approaches the marginal distribution of ε as the number of observations increases and the step length, Δt, is kept fixed. Using this notation for the random fields involved in the current problem, the functions used in the method of moments are,

$$\boldsymbol{I}(|\gamma\, \varepsilon'_C + \varepsilon'_G| > K)$$

and

$$\begin{bmatrix} \alpha'_L \\ \beta'_L \\ \rho'_L \end{bmatrix} [\,\alpha'_L,\, \beta'_L,\, \rho'_L\,] \boldsymbol{I}(|\gamma\varepsilon'_C + \varepsilon'_G| < K),$$

with γ being the global scale for the Cauchy process; K being a fixed constant; $\boldsymbol{I}(\cdot)$ being an indicator function for an event; ε'_C being as for Expression (4); and ε'_G being a linear combination of the Gaussian processes, ε'_1, ε'_2 and ε'_3, see Expression (4). In the current setting, $K = 0.0495$ and

$$\gamma\, \varepsilon'_C + \varepsilon'_G = 0.60\, \alpha'_L + 0.32\, \beta'_L + 0.54\, \rho'_L$$

The estimates must be solved numerically. Since the main focus is not on these parameters, approximate values for the population averages are computed, by using the approximations

$$P(|\gamma\, \varepsilon'_C + \varepsilon'_G| > K) \approx P(|\varepsilon'_C| > K/\gamma)$$

and

$$P(|\varepsilon'_C| > K/\gamma) \approx \frac{2\gamma}{\pi K}.$$

The approximations improves as K increases.

E Simplifications by using the Fourier transform

The expressions stated in the previous sections are valid for more general models than the particular that is specified. In the specified model, several expressions

simplify since the Fourier transform simultaneously diagonalize the stationary operators considered. To use the simplified formulas the convolutions must be defined cyclic. This is done by tapering the data and extend the vectors by zero padding to avoid boundary problems.

Let $\boldsymbol{X}^T$ denote the orthogonal Fourier transform, and $\boldsymbol{X}$ denote its inverse. Now introduce the relations,

$$\begin{aligned}
\boldsymbol{S}_k &= \boldsymbol{X}\boldsymbol{\Lambda}_k^s\boldsymbol{X}^T\,, \quad k \in \{1,2,...,m\}\,,\\
\boldsymbol{\Phi}_j &= \boldsymbol{X}\boldsymbol{\Lambda}^j\boldsymbol{X}^T\,, \quad j \in \{C,\,1,\,2,\,3\,\}\,,\\
\boldsymbol{\Sigma}_E &= \boldsymbol{\Sigma}^\theta \otimes \boldsymbol{\Sigma}^t\,,\\
\boldsymbol{\Sigma}^t &= \boldsymbol{X}\boldsymbol{\Lambda}^t\boldsymbol{X}^T\,,\\
\boldsymbol{X}^{\otimes m} &= \boldsymbol{I}_{m\times m} \otimes \boldsymbol{X}\,,\\
\boldsymbol{X}^{T\otimes m} &= \boldsymbol{I}_{m\times m} \otimes \boldsymbol{X}^T\,,
\end{aligned}$$

with the $\boldsymbol{\Lambda}$'s being diagonal matrices; $\otimes$ being the Kronecker tensor product; and

$$\boldsymbol{\Sigma}^\theta = \begin{bmatrix} \sigma^\theta_{11} & \dots & \sigma^\theta_{1m} \\ \vdots & \ddots & \vdots \\ \sigma^\theta_{m1} & \dots & \sigma^\theta_{mm} \end{bmatrix}\,.$$

According to these identities;

$$\boldsymbol{\Sigma}_N = \sigma^2\boldsymbol{K}\boldsymbol{K}^T + \boldsymbol{\Sigma}_E = \boldsymbol{X}^{\otimes m}\boldsymbol{D}_N\boldsymbol{X}^{T\otimes m}$$

with

$$\boldsymbol{D}_N = \begin{bmatrix} \boldsymbol{D}_{11} & \dots & \boldsymbol{D}_{1m} \\ \vdots & \ddots & \vdots \\ \boldsymbol{D}_{m1} & \dots & \boldsymbol{D}_{mm} \end{bmatrix}\,,$$

with $\boldsymbol{D}_{kl}$ being diagonal matrices having the form:

$$\boldsymbol{D}_{kl} = \sigma^\theta_{kl}\boldsymbol{\Lambda}^t + \sigma^2\sum_{j=1}^{3} a^j_k a^j_l \boldsymbol{\Lambda}^s_k\boldsymbol{\Lambda}^s_l(\boldsymbol{\Lambda}^j)^2.$$

The inverse of $\boldsymbol{\Sigma}_N$ can be calculated by solving a $m\times m$ system for each Fourier component separately. When $\boldsymbol{\Sigma}_E$ have a small white noise component the inversion is stable. The inverse of $\boldsymbol{D}_N$ have the same structure as $\boldsymbol{D}_N$,

$$\boldsymbol{D}_N{}^{-1} = \begin{bmatrix} \boldsymbol{D}^{11} & \dots & \boldsymbol{D}^{1m} \\ \vdots & \ddots & \vdots \\ \boldsymbol{D}^{m1} & \dots & \boldsymbol{D}^{mm} \end{bmatrix}\,.$$

with $\boldsymbol{D}^{kl}$ being diagonal matrices.

For the sampling the Cauchy seed, σ_{l^*} and $\boldsymbol{\mu}_{l^*}$, need to be computed. First simplify the matrix $\boldsymbol{K}_{\mathrm{C}}^T \boldsymbol{\Sigma}_N^{-1} \boldsymbol{K}_{\mathrm{C}}$ using the notation above,

$$\boldsymbol{K}_{\mathrm{C}}^T \boldsymbol{\Sigma}_N^{-1} \boldsymbol{K}_{\mathrm{C}} = \boldsymbol{X}^T \boldsymbol{\Lambda}_H^2 \boldsymbol{X}^T$$

with

$$\boldsymbol{\Lambda}_H^2 = \sum_{k=1}^{m} \sum_{l=1}^{m} a_k^C a_l^C \boldsymbol{\Lambda}_k^s \boldsymbol{\Lambda}_l^s \boldsymbol{D}^{kl} (\boldsymbol{\Lambda}^C)^2,$$

hence:

$$\sigma_{l^*} = \| \boldsymbol{\Lambda}_H \boldsymbol{X}^T \boldsymbol{v}^{l^*} \|_2^{-1},$$

$$\boldsymbol{\mu}_{l^*}^T \boldsymbol{K}_{\mathrm{C}} \boldsymbol{c}^{\mathrm{old}} = \sigma_{l^*}^2 (\boldsymbol{X}^T \boldsymbol{v}^{l^*})^T \boldsymbol{\Lambda}_H^2 \boldsymbol{X}^T \boldsymbol{c}^{\mathrm{old}},$$

$$\boldsymbol{\mu}_{l^*}^T \boldsymbol{d} = \sigma_{l^*}^2 (\boldsymbol{X}^T \boldsymbol{v}^{l^*})^T \left(\sum_{k=1}^{m} \sum_{l=1}^{m} \frac{a_k^C \boldsymbol{\Lambda}_k^s \boldsymbol{D}^{kl} \boldsymbol{\Lambda}^C}{\boldsymbol{\Lambda}_H} \boldsymbol{X}^T \boldsymbol{d}_l \right),$$

with $\boldsymbol{d}_l$ being the discretized traces at angle θ_l.

For the normal seed computations, the important quantities are $\boldsymbol{\mu}_{\mathrm{w}}$ and $\boldsymbol{\Sigma}_{\mathrm{w}}$. These are split into the components corresponding to each of the normal seeds, $\boldsymbol{W}_j;\ j \in \{1, 2, 3\}$,

$$\boldsymbol{\mu}_{\mathrm{w}} = \begin{bmatrix} \boldsymbol{\mu}_1^{\mathrm{w}} \\ \boldsymbol{\mu}_2^{\mathrm{w}} \\ \boldsymbol{\mu}_3^{\mathrm{w}} \end{bmatrix}$$

and,

$$\boldsymbol{\Sigma}_{\mathrm{w}} = \begin{bmatrix} \boldsymbol{\Sigma}_{11}^{\mathrm{w}} & \boldsymbol{\Sigma}_{12}^{\mathrm{w}} & \boldsymbol{\Sigma}_{13}^{\mathrm{w}} \\ \boldsymbol{\Sigma}_{21}^{\mathrm{w}} & \boldsymbol{\Sigma}_{22}^{\mathrm{w}} & \boldsymbol{\Sigma}_{23}^{\mathrm{w}} \\ \boldsymbol{\Sigma}_{31}^{\mathrm{w}} & \boldsymbol{\Sigma}_{32}^{\mathrm{w}} & \boldsymbol{\Sigma}_{33}^{\mathrm{w}} \end{bmatrix}.$$

By the above notation,

$$\boldsymbol{\mu}_j^{\mathrm{w}} = \sigma^2 \boldsymbol{X} \sum_{k=1}^{m} \sum_{l=1}^{m} a_k^j \boldsymbol{D}^{kl} \boldsymbol{\Lambda}^j \boldsymbol{\Lambda}_k^s \boldsymbol{X}^T (\boldsymbol{d}_l - (\boldsymbol{K}_{\mathrm{C}} \boldsymbol{c}^{\mathrm{old}})_l),$$

with $(\boldsymbol{K}_{\mathrm{C}} \boldsymbol{c}^{\mathrm{old}})_l$ being the vector containing the, $(l-1)n+1$, $(l-1)n+2, ..., ln$, components of $\boldsymbol{K}_{\mathrm{C}} \boldsymbol{c}^{\mathrm{old}}$, further

$$\boldsymbol{\Sigma}_{ij}^{\mathrm{w}} = \sigma^2 \delta_{ij} \cdot \boldsymbol{I} - \sigma^4 \boldsymbol{X} \left(\sum_{k=1}^{m} \sum_{l=1}^{m} a_k^j a_l^i \boldsymbol{D}^{kl} \boldsymbol{\Lambda}^j \boldsymbol{\Lambda}^i \boldsymbol{\Lambda}_k^s \boldsymbol{\Lambda}_l^s \right) \boldsymbol{X}^T,$$

with $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ if $i \neq j$,

The random seeds can hence be sampled by sampling the frequencies of $\boldsymbol{W}_1, \boldsymbol{W}_2, \boldsymbol{W}_3$ simultaneously. For each of the n frequencies, a 3×3 matrix must be factored.

Tables and figures

material parameter	level	scale ε_C	scale ε_1	scale ε_1	scale ε_3
α	7.9941	0.0059	0.0183	-0.0014	0.0004
β	7.2651	0.0059	-0.0014	0.0303	-0.0014
ρ	7.7629	0.0009	0.0004	-0.0014	0.0106

Table 1: Level and scale parameters for the Cauchy model. The table relates to Expression (4).

material parameter	level	scale ε_C	scale ε_1	scale ε_1	scale ε_3
α	7.9941	0.0000	0.0319	0.0102	0.0032
β	7.2651	0.0000	0.0102	0.0449	0.0007
ρ	7.7629	0.0000	0.0032	0.0007	0.0120

Table 2: Level and scale parameters for the pure Gaussian model. The table relates to Expression (4).

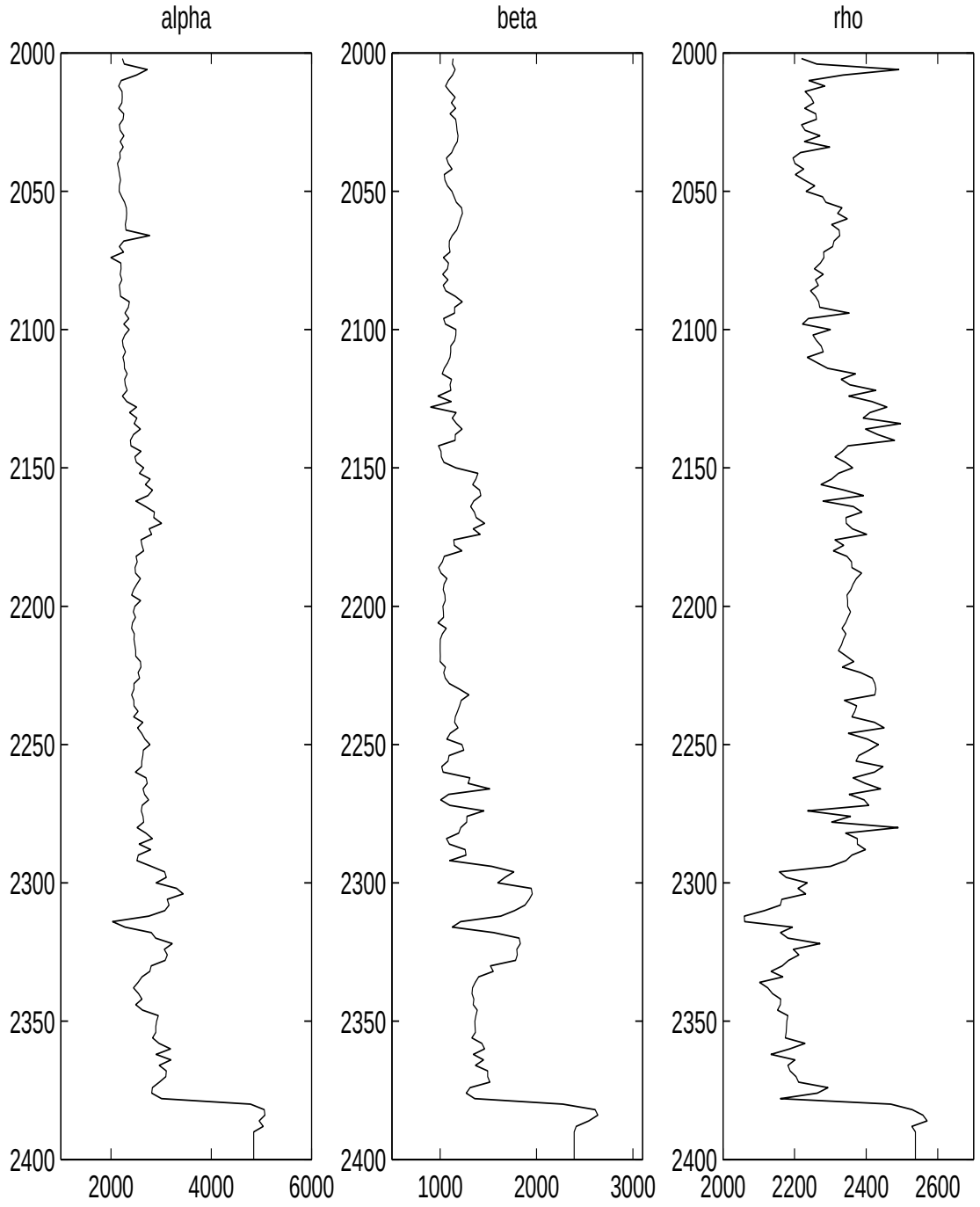

Figure 1: Well logs. The material parameters observed in a well at the Sleipner Øst Field. The observed depth profile is converted to a time profile. The drilling stopped at 2390 ms below this depth the value is fixed at a constant level.

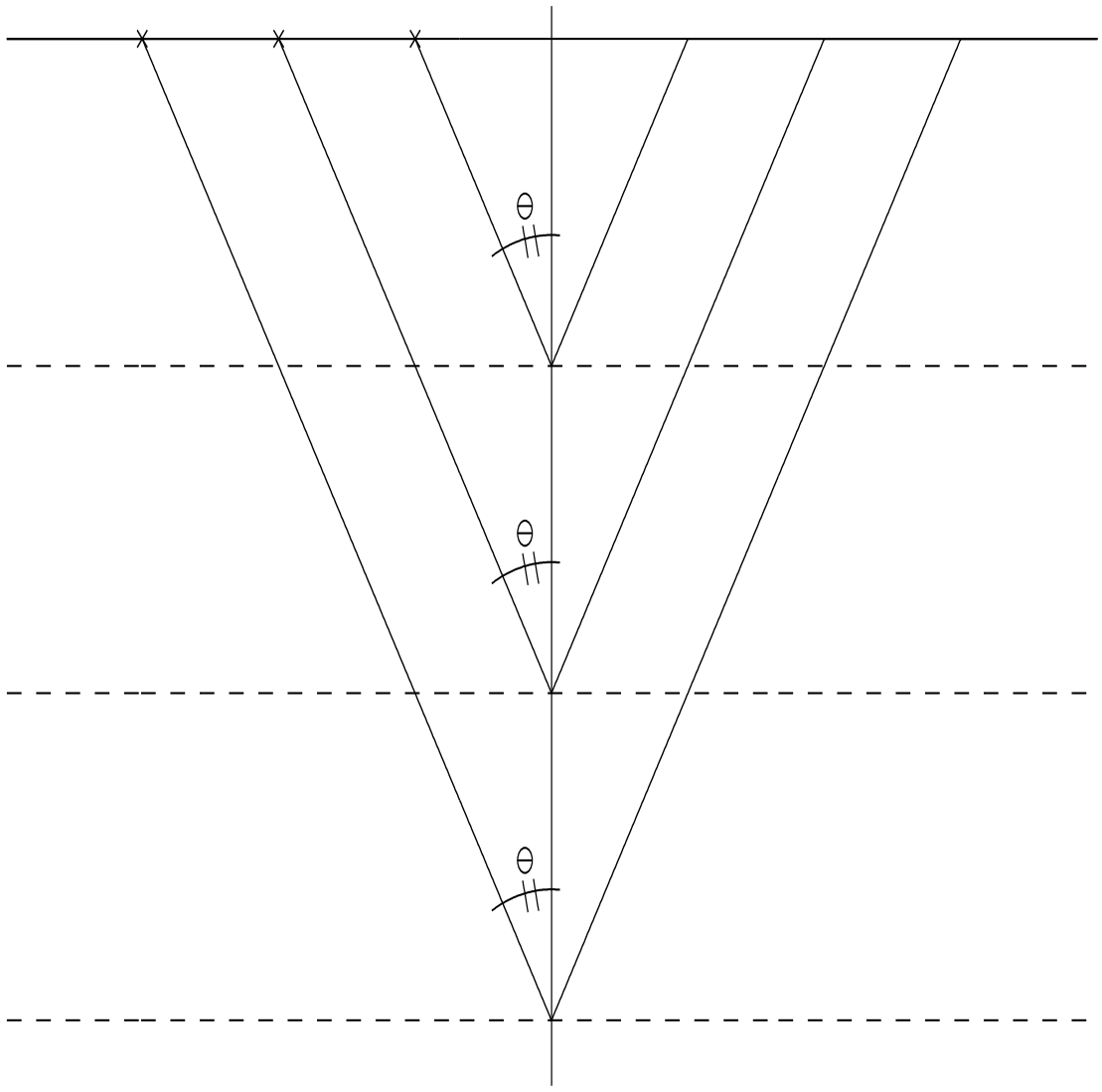

Figure 2: Angle gather-common angle. The ray paths all have a common angle to the vertical line, and hence a common angle in the angle gather

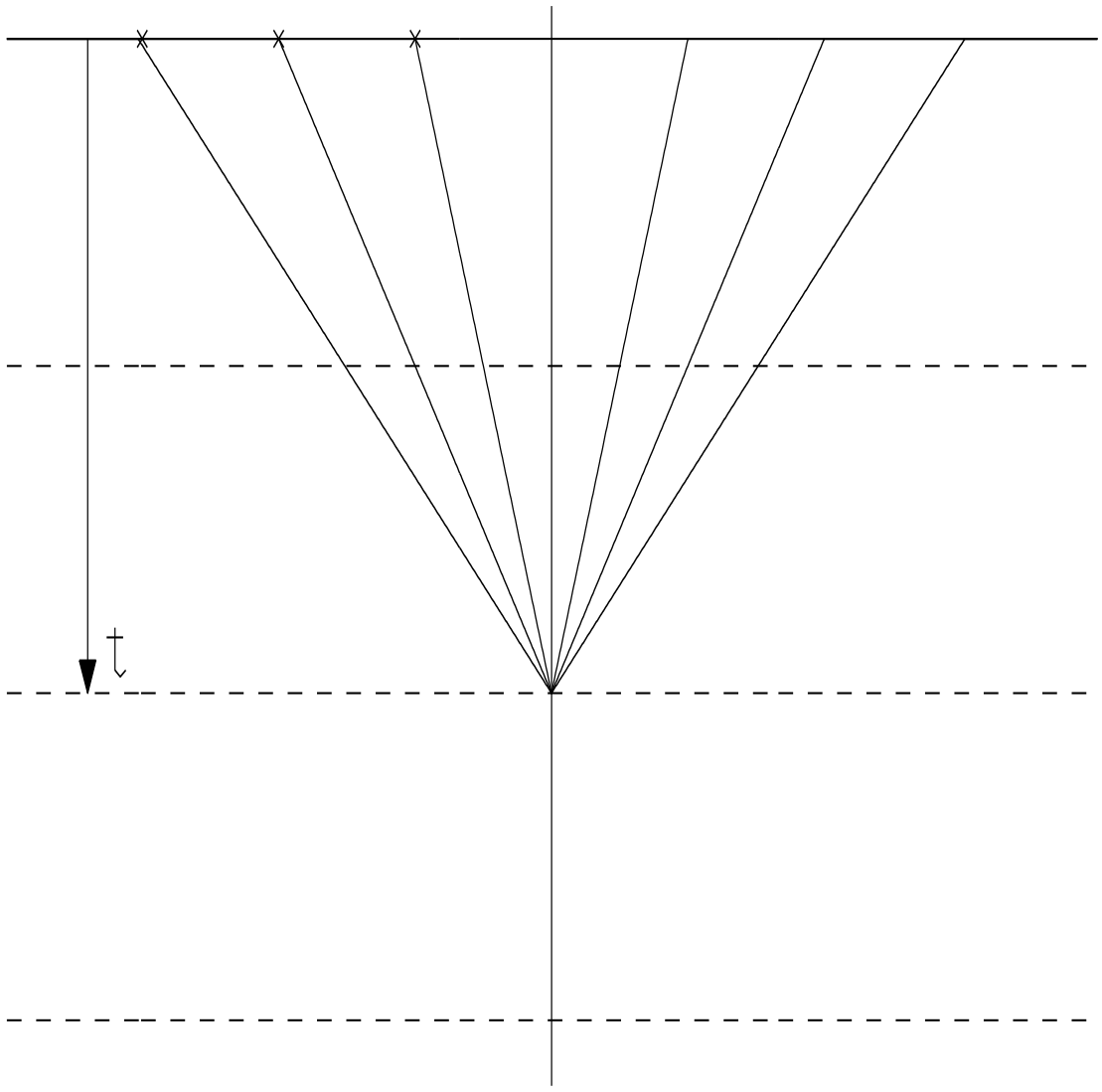

Figure 3: Angle gather-common time. The ray paths all have a common depth of reflection, and hence the same time reference in the angle gather.

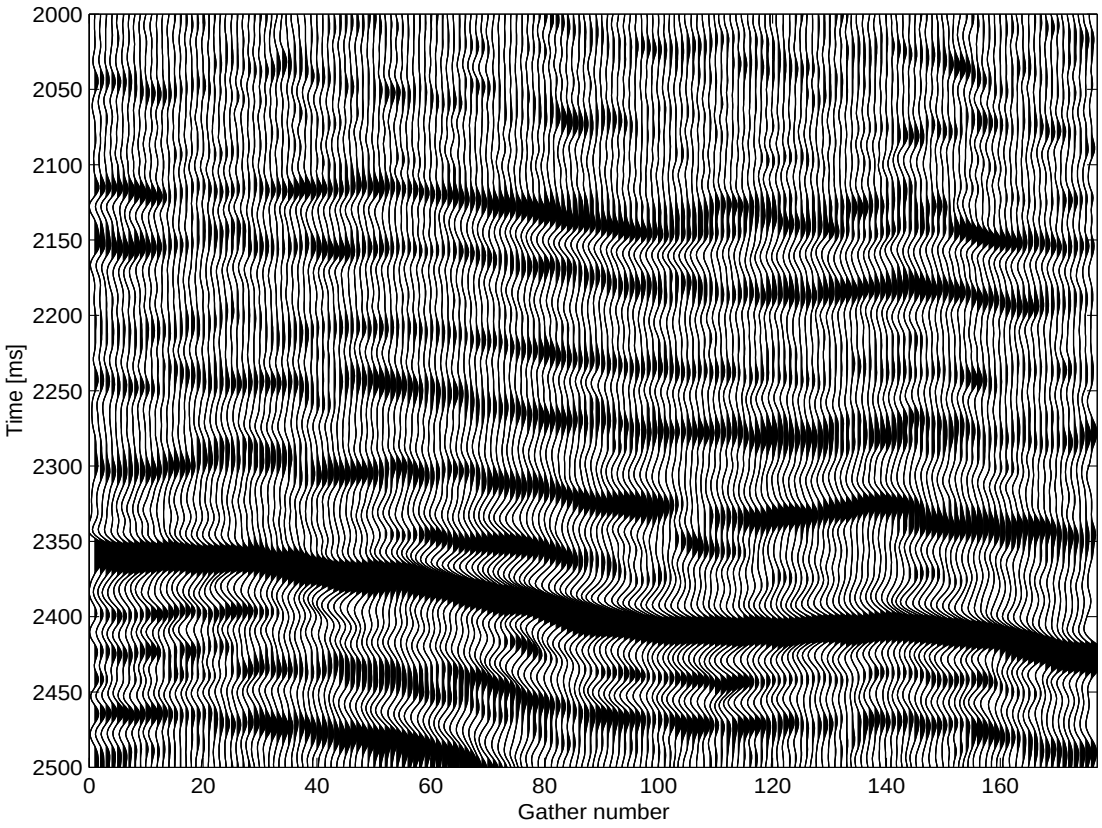

Figure 4: Stack section for the seismic inline. The seismic inline is observed in a marine seismic survey above the Sleipner Øst Field.

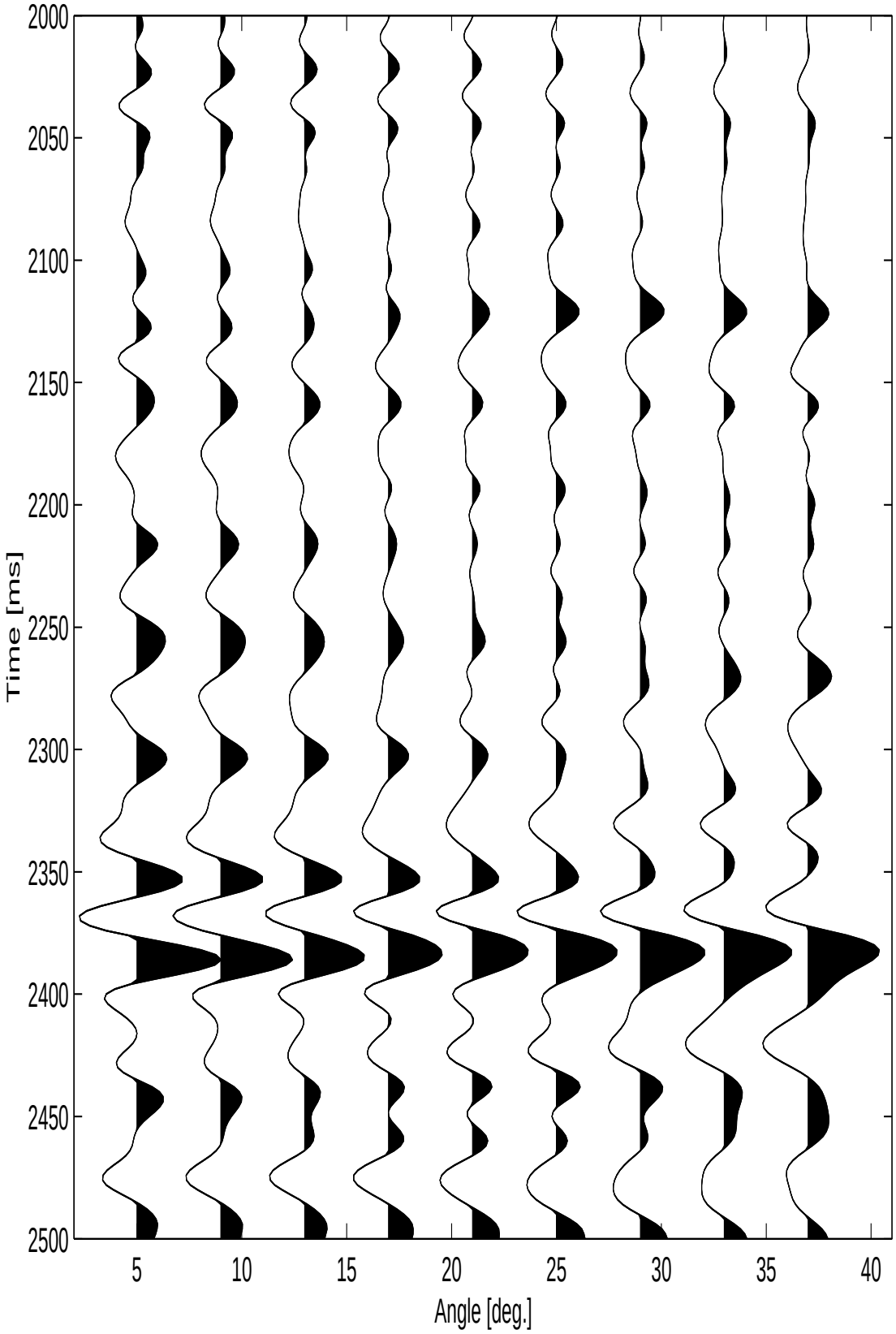

Figure 5: Angle gather number 67. The seismic traces in this angle gather is recorded in the well location.

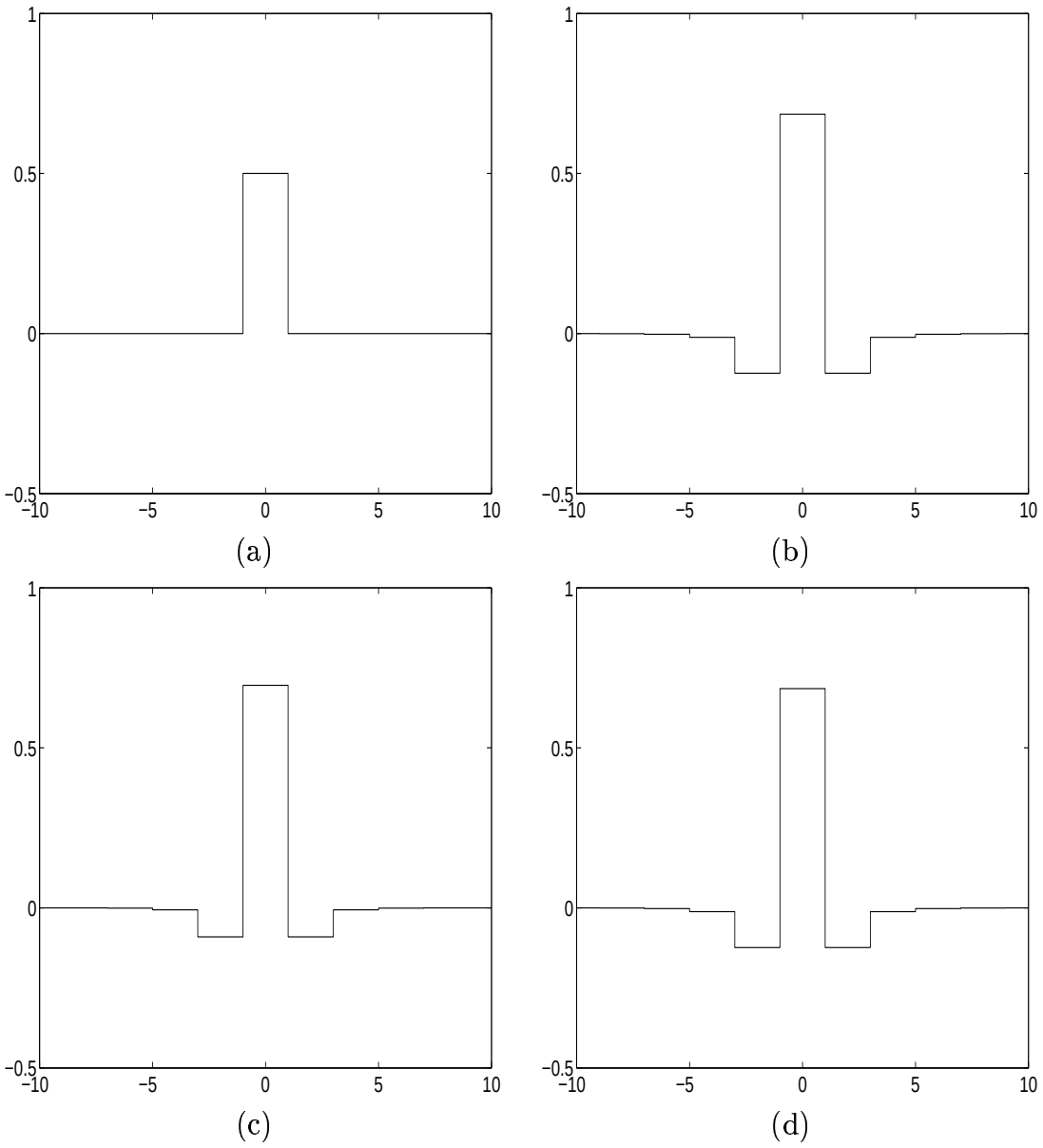

Figure 6: Smoothing kernels for the Cauchy model. The kernels are relates to the random components of Expression (4). The functions shown are, $\phi'_C(t)$ (a); $\phi'_1(t)$ (b); $\phi'_2(t)$ (c) and $\phi'_3(t)$ (d), respectively.

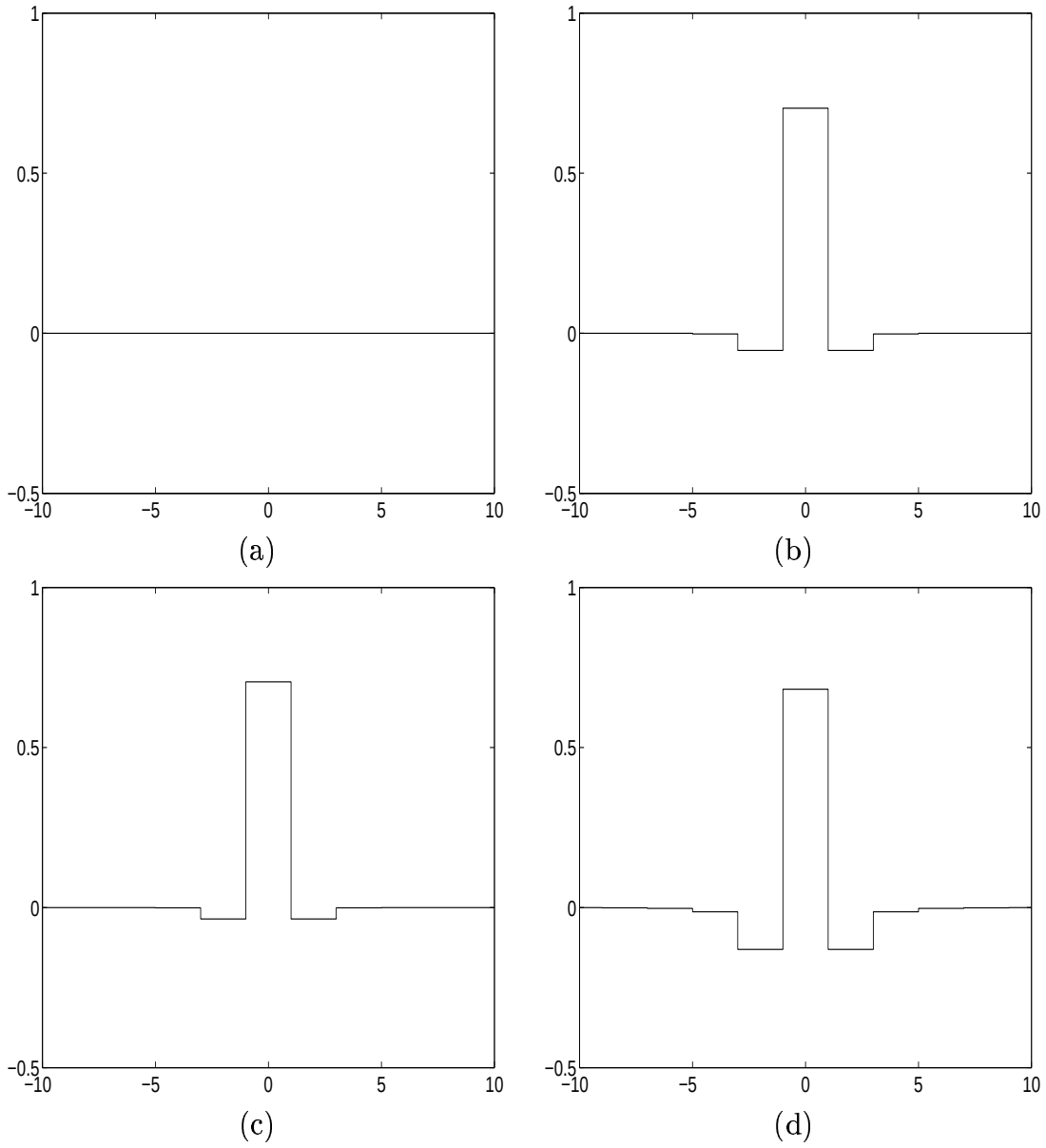

Figure 7: Smoothing kernels for the pure Gaussian model. The kernels are relates to the random components of Expression (4). The functions shown are, $\phi'_C(t)$ (a); $\phi'_1(t)$ (b); $\phi'_2(t)$ (c) and $\phi'_3(t)$ (d), respectively.

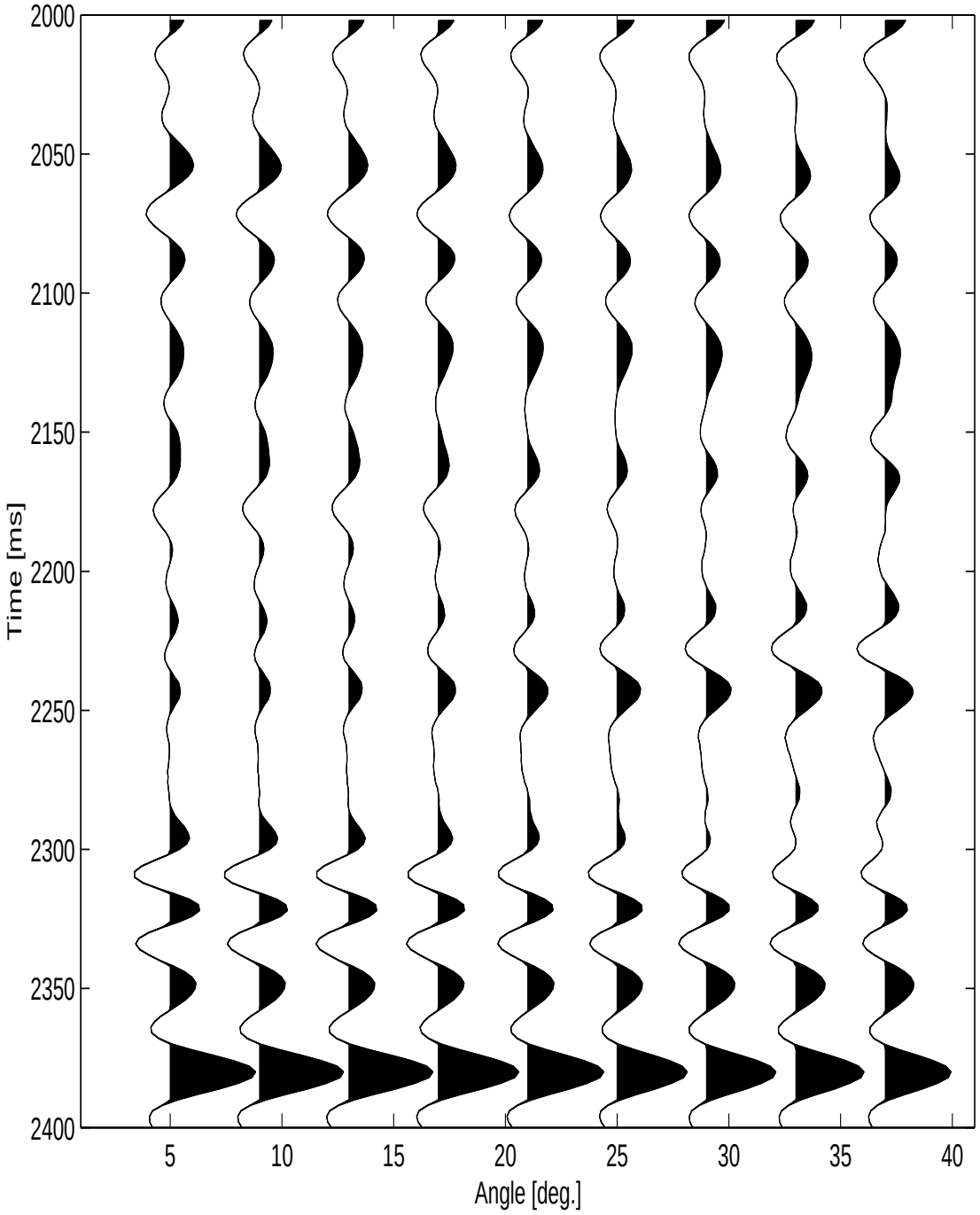

Figure 8: Synthetic observations. The CDP gather inverted in the example in Section 6. The observations are generated by the linearized model and errors where added according to the likelihood.

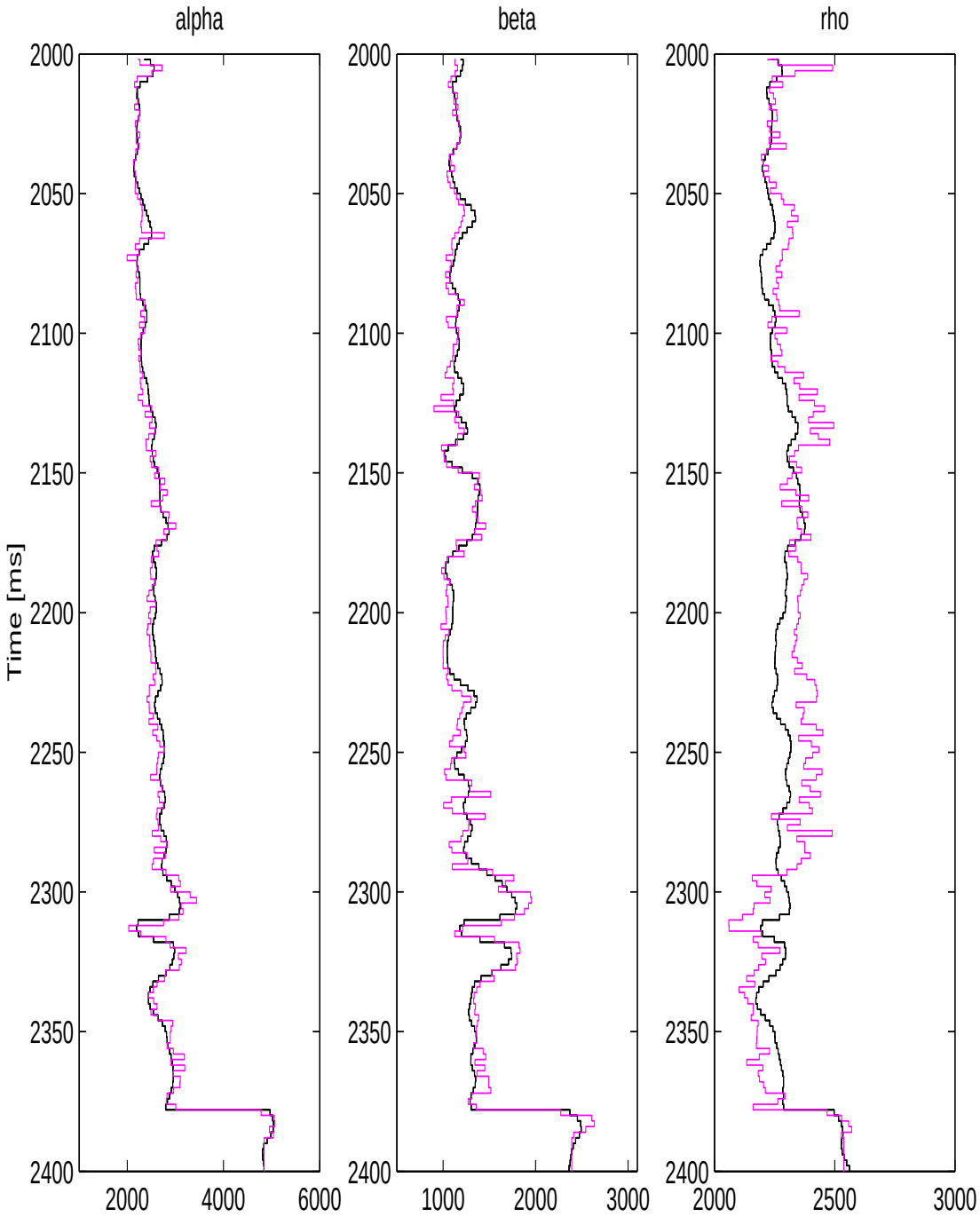

Figure 9: Inversion results for the Cauchy model. The median and is plotted together with the true parameter values.

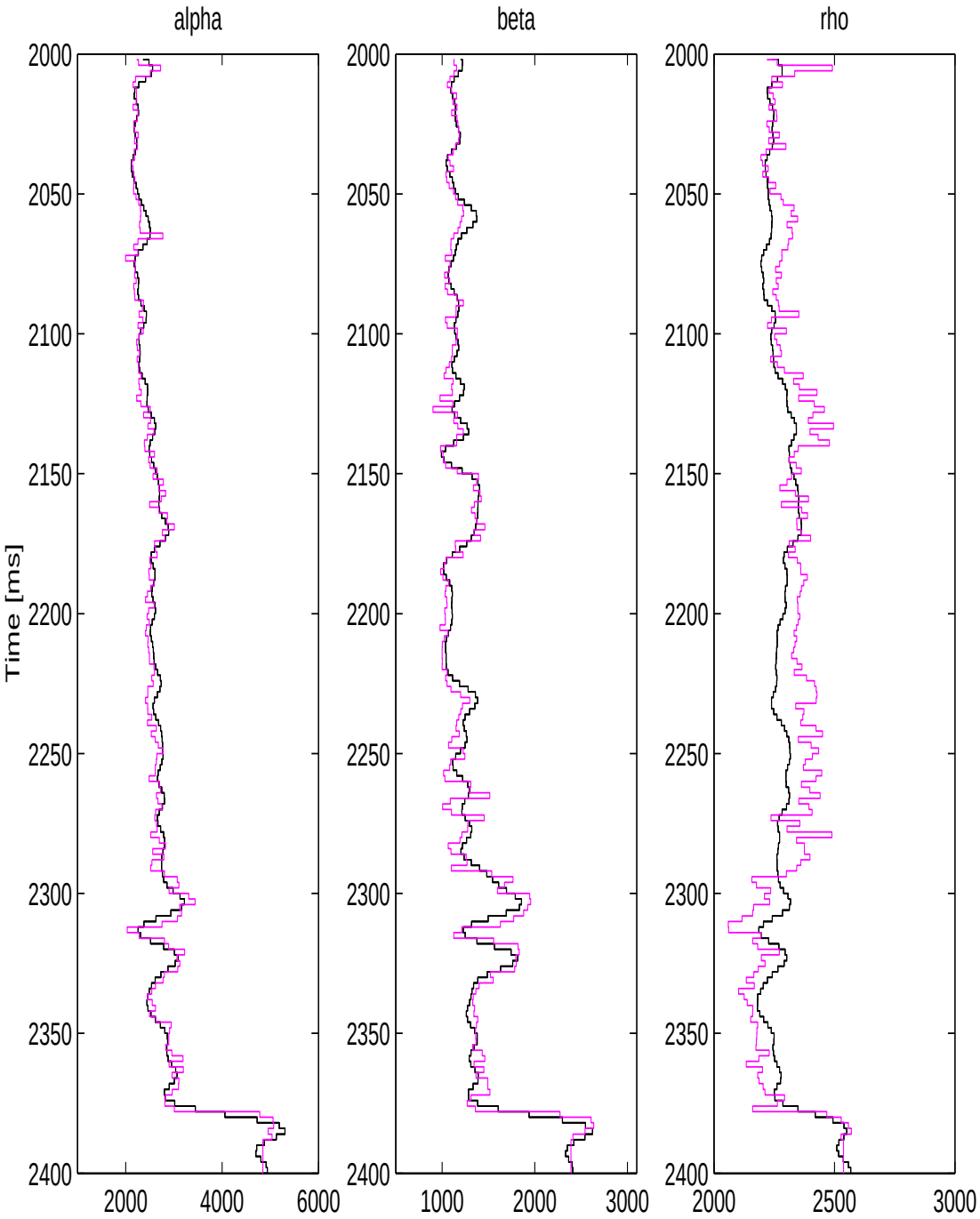

Figure 10: Inversion results for the pure Gaussian model. The median is plotted together with the true parameter values.

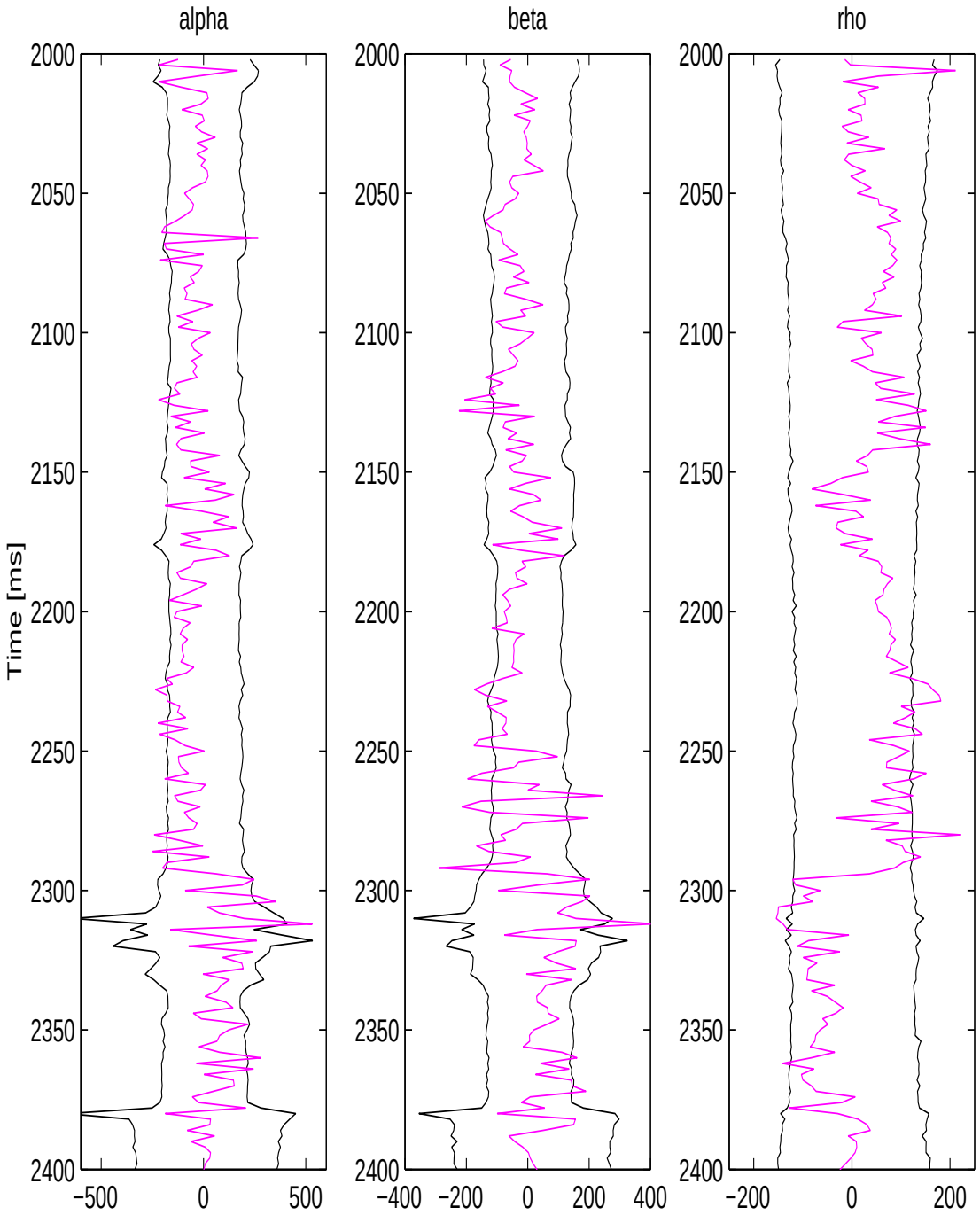

Figure 11: Errors for the Cauchy model. The 90% credibility interval for the error predicted by the simulations displayed together with the actual error.

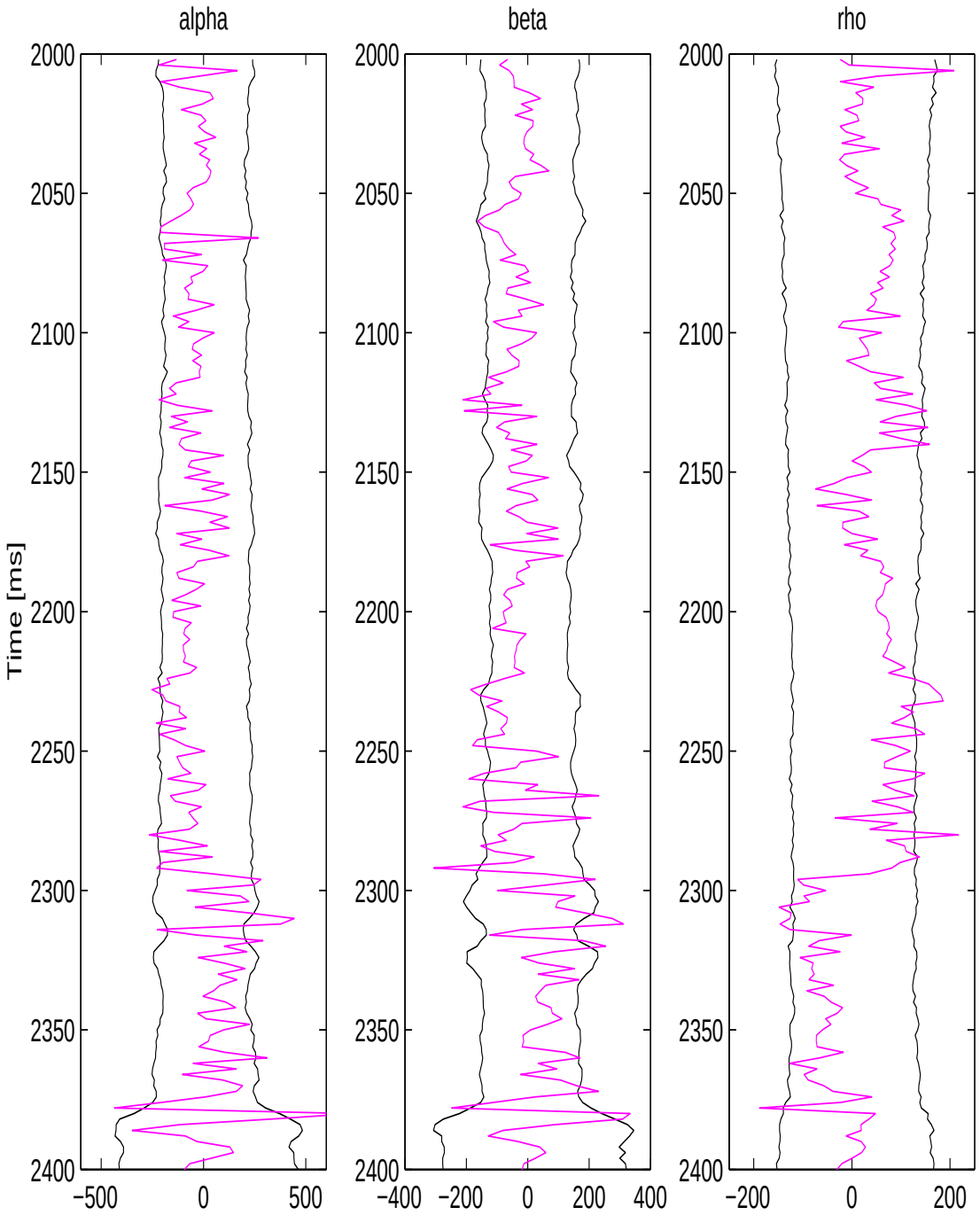

Figure 12: Errors for the pure Gaussian model. The 90% credibility interval for the error predicted by the simulations displayed together with the actual error.

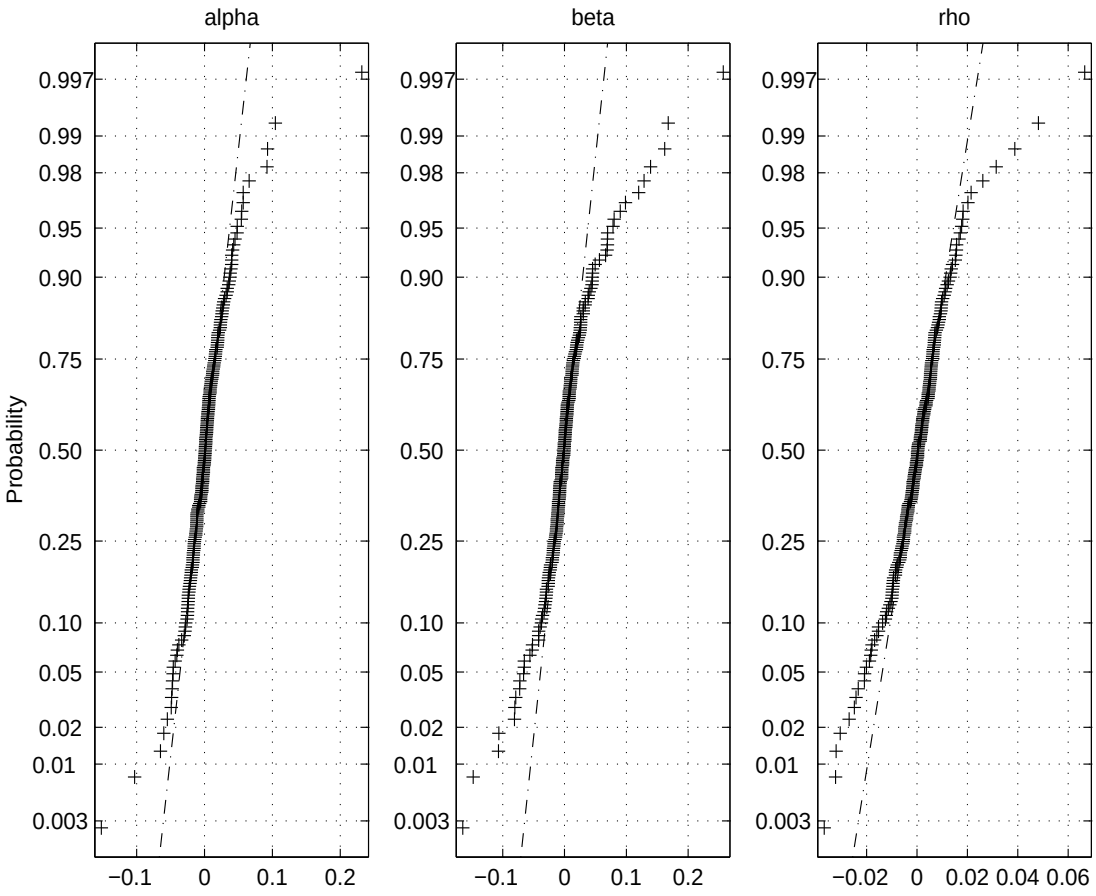

Figure 13: Normal plots for the derivative of the logarithm of the material parameters.

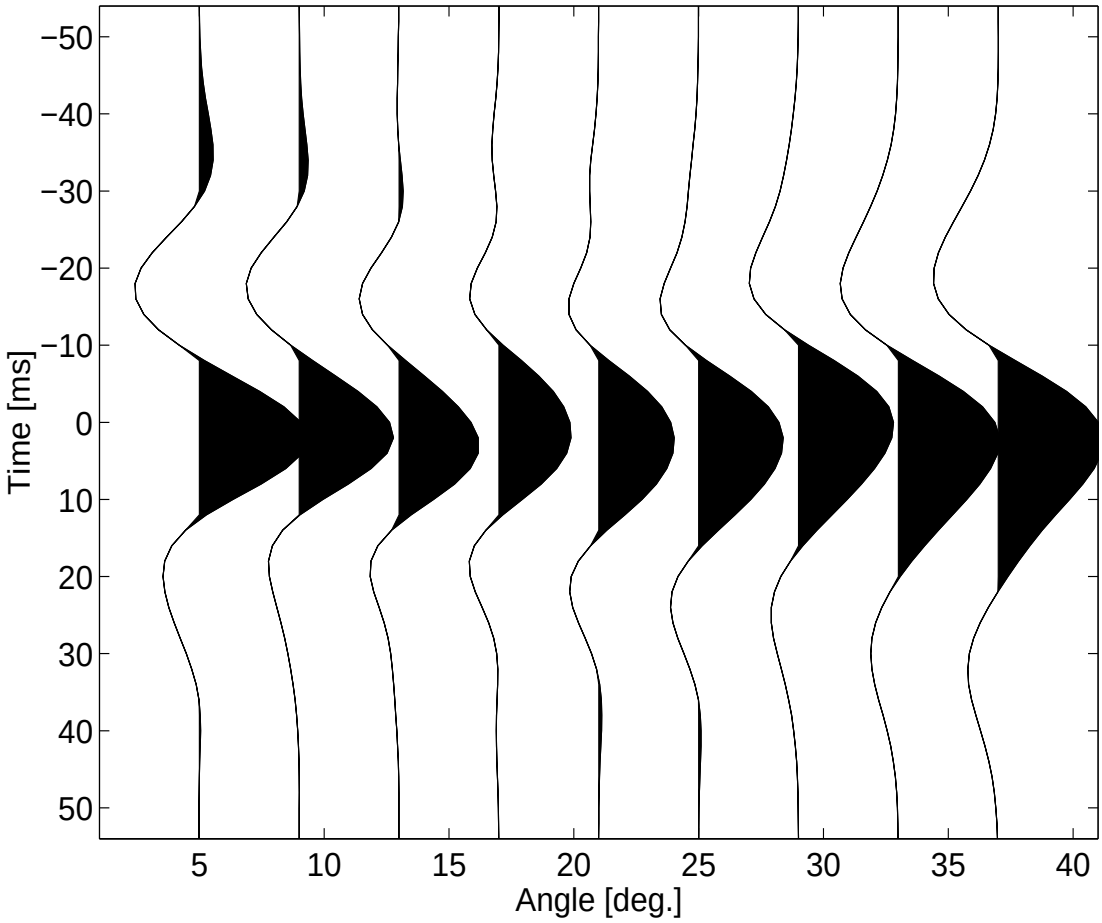

Figure 14: The wavelets for the inversion. The wavelets where estimated from the well log displayed in Figure 1 and the angle gather displayed in Figure 5.

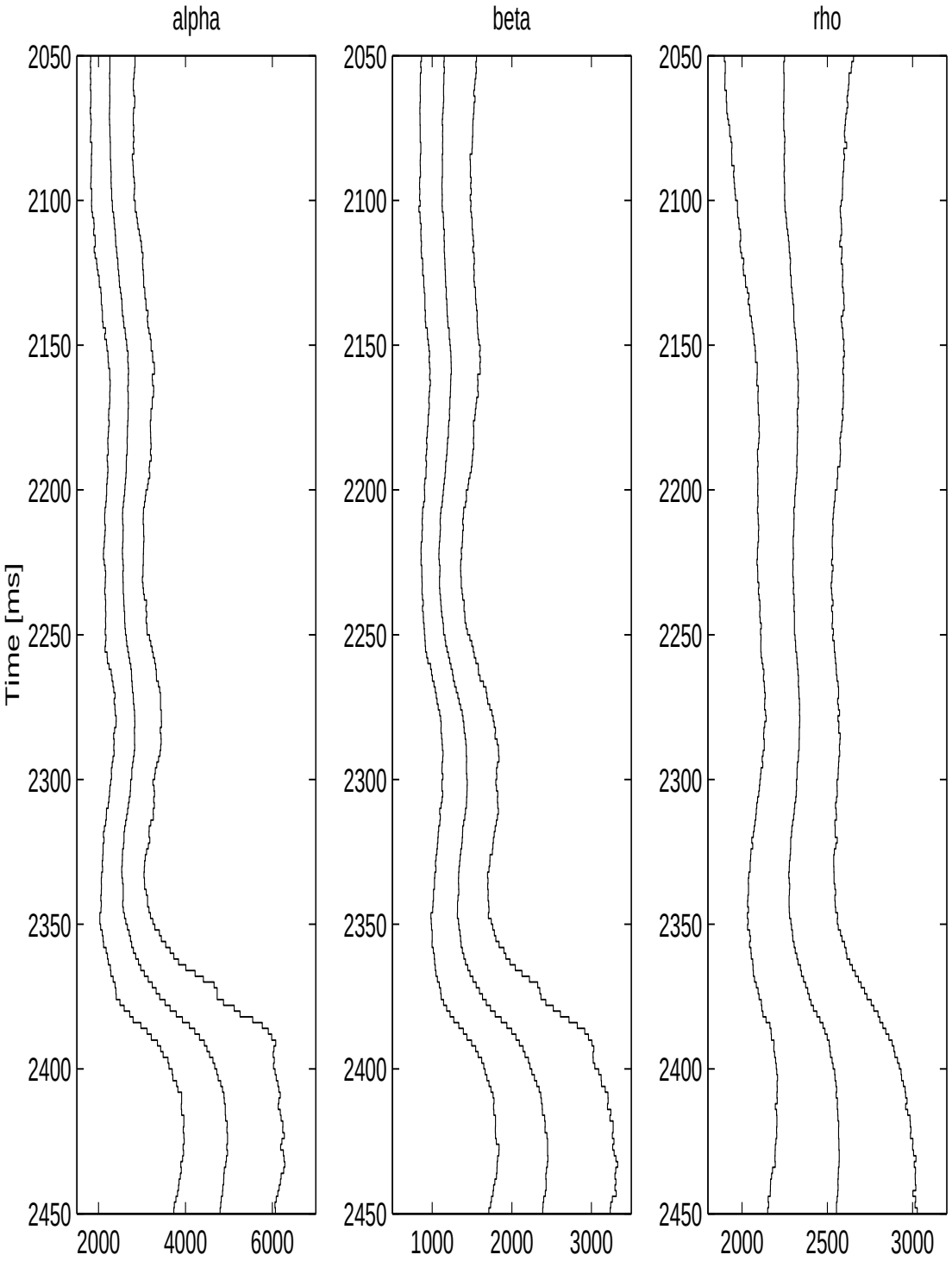

Figure 15: The prior distribution gather 67. The prior median and pointwise 90% credibility interval.

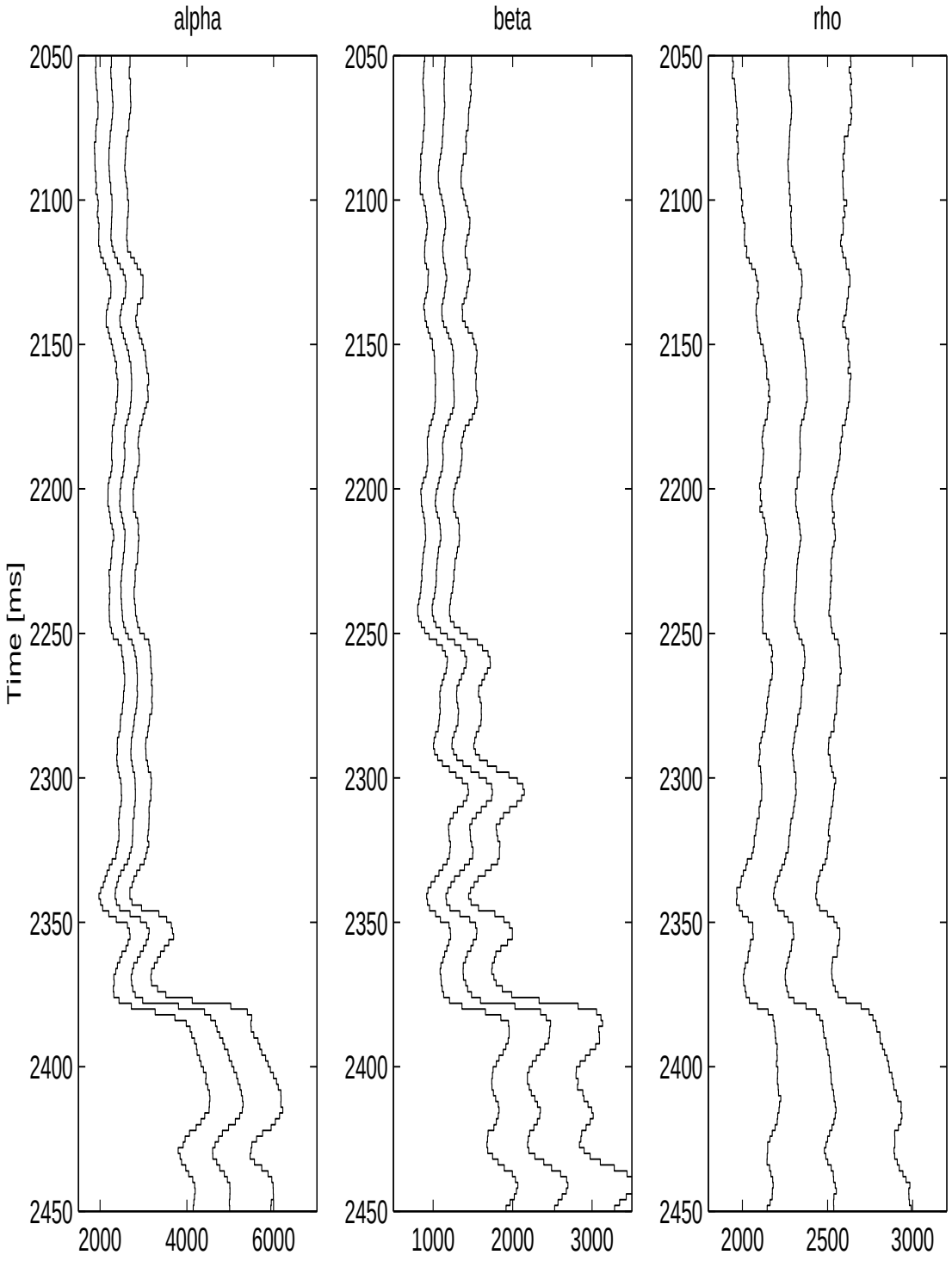

Figure 16: The posterior distribution gather 67. The posterior median and pointwise 90% credibility interval.

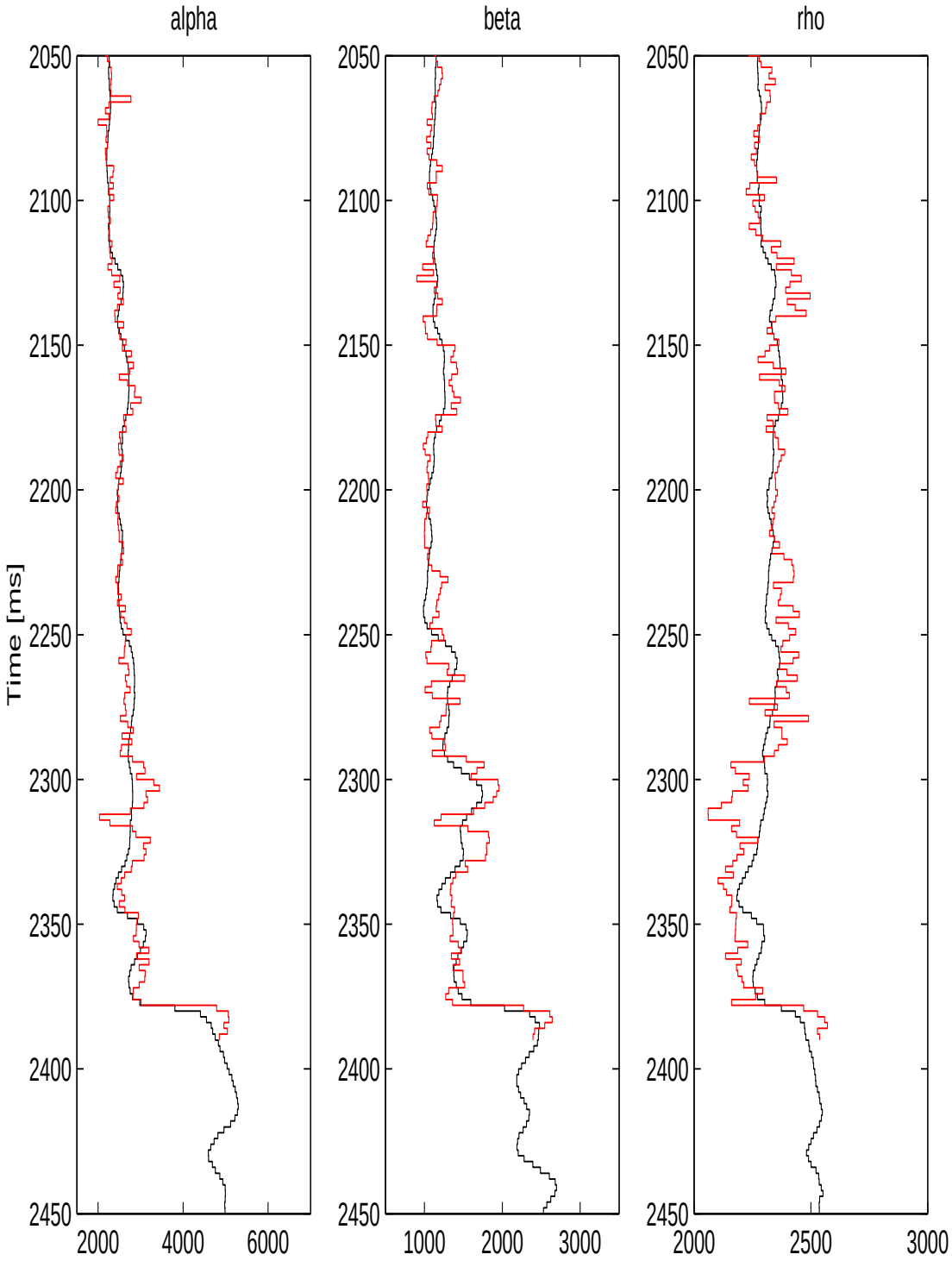

Figure 17: The inversion results for Sleipner Øst gather number 67. The figure display the inverted path together with the values observed in the well in red.

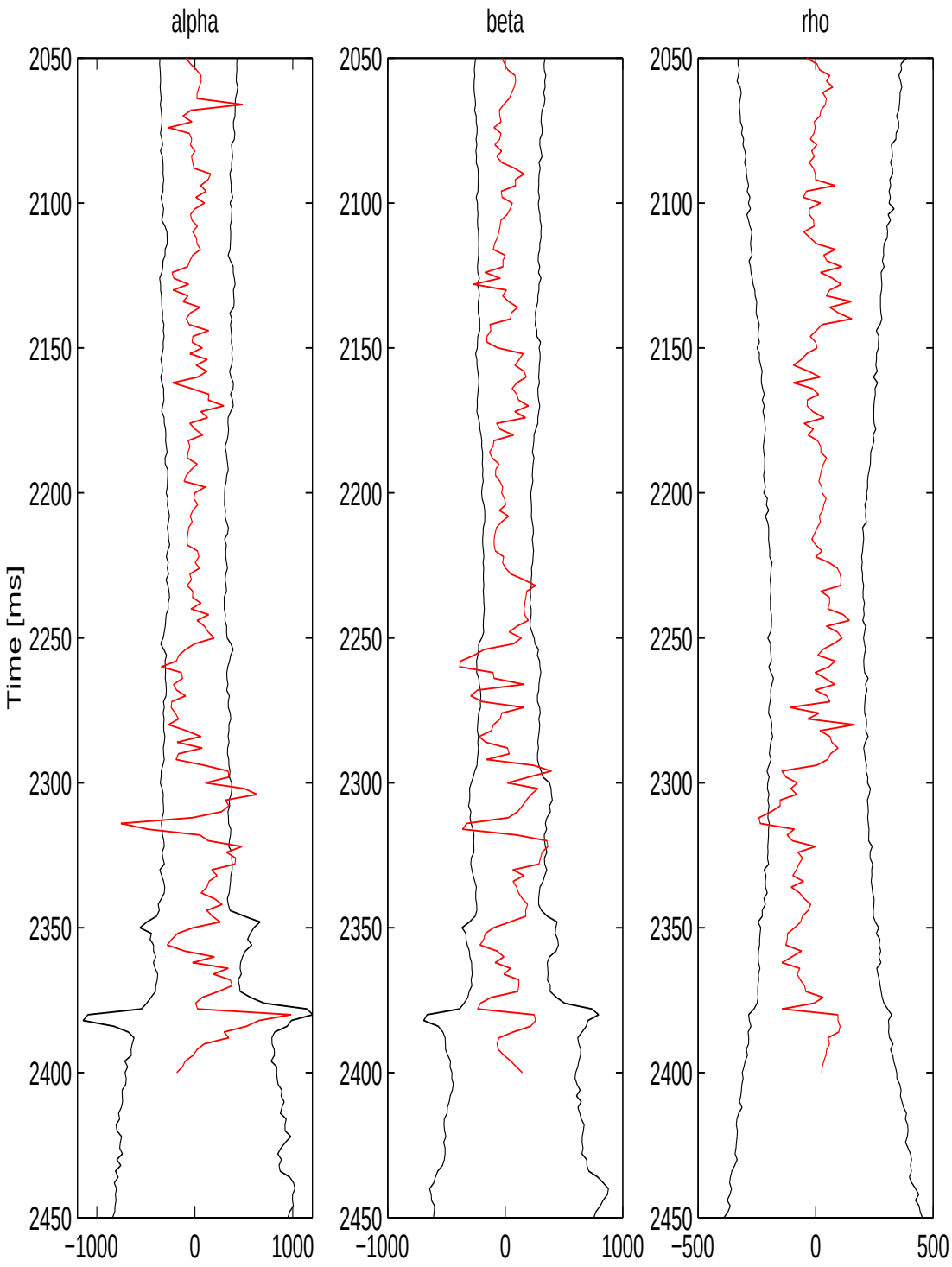

Figure 18: Errors for for Sleipner Øst gather number 67. The 90% credibility interval for the error predicted by the simulations displayed together with the actual error.

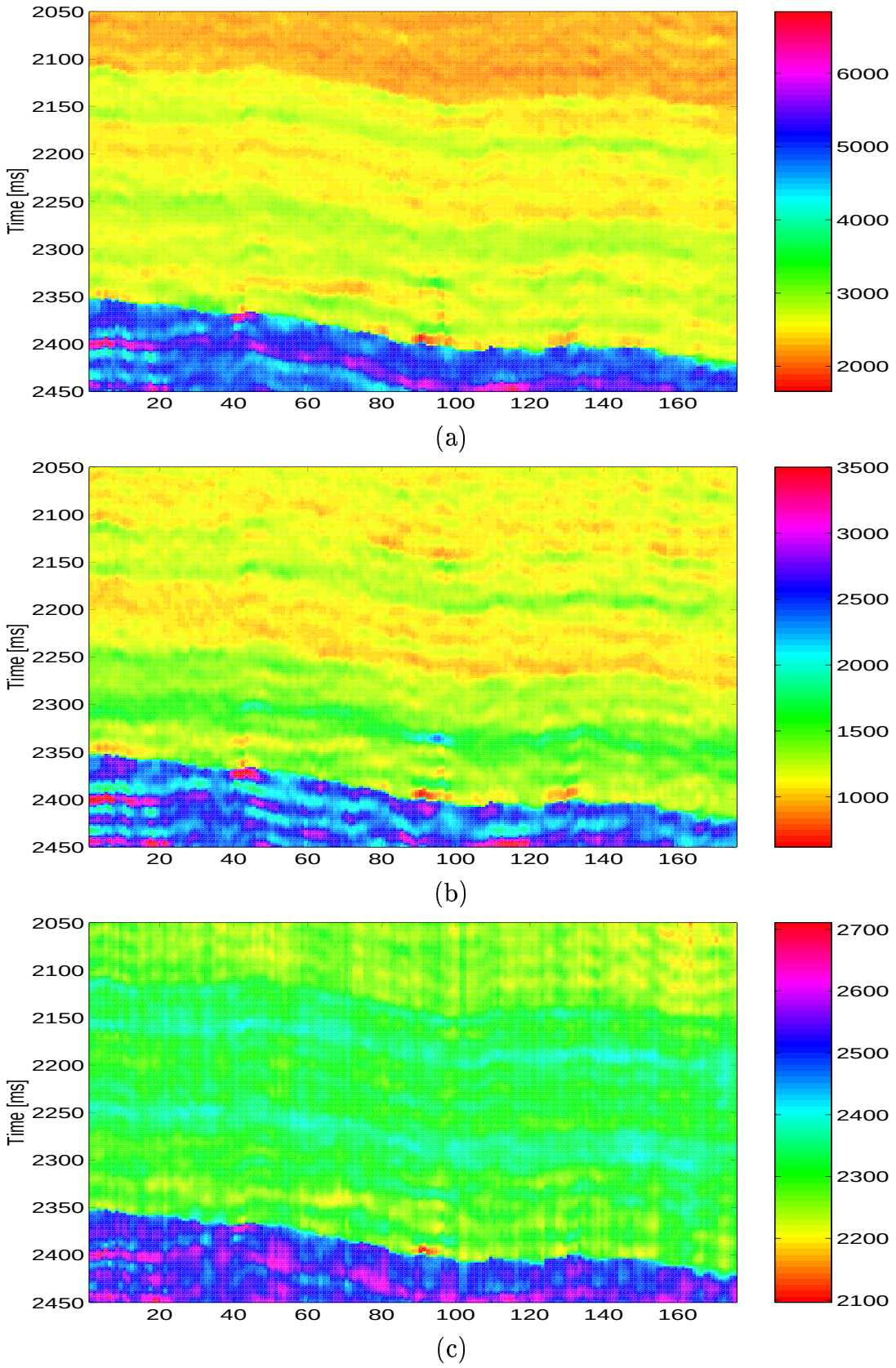

Figure 19: The inversion results for Sleipner Øst data. The estimated values being the posterior medians for (a) α; (b) β; and (c) ρ, respectively.

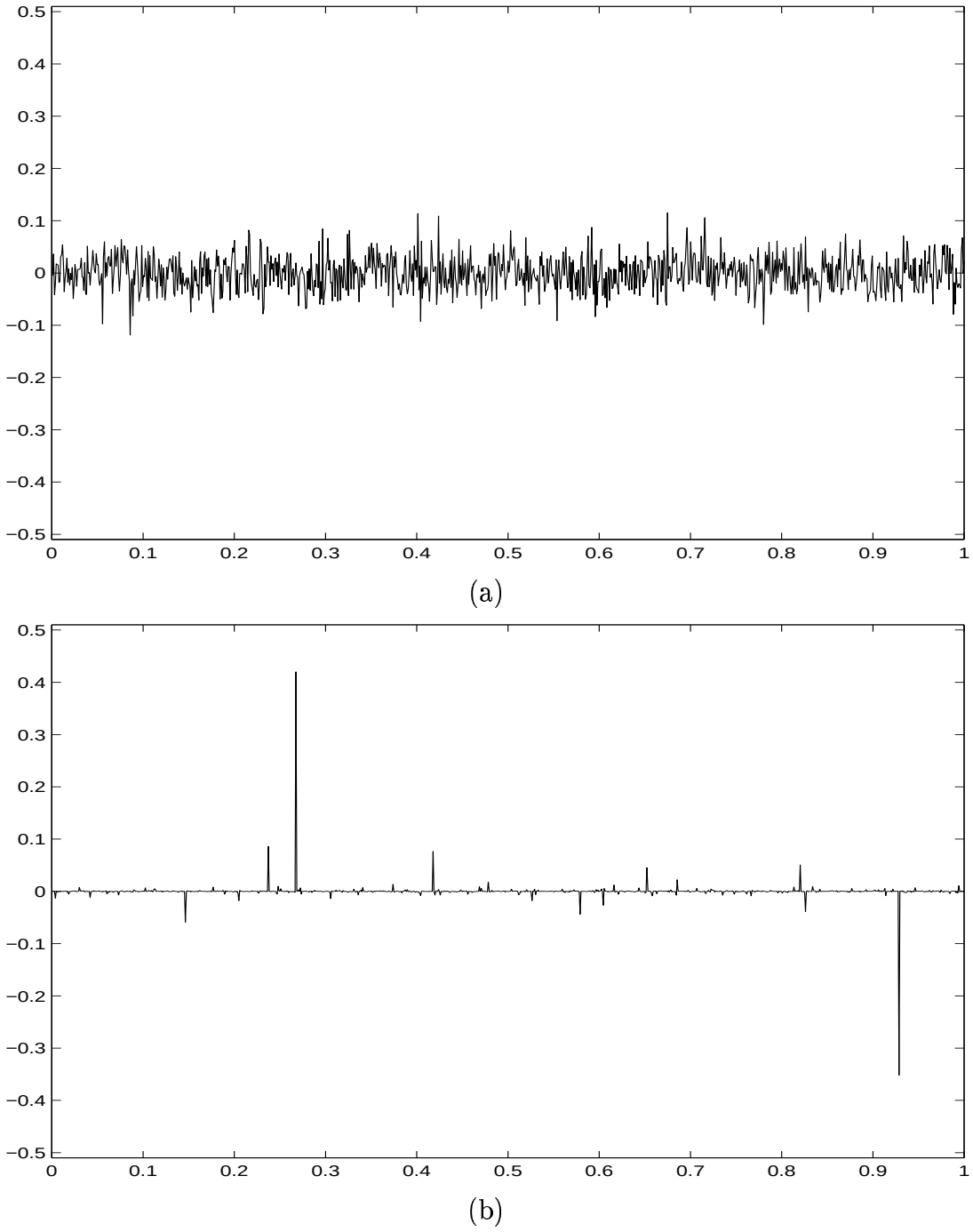

Figure 20: Comparison of the Winer measure and the Cauchy measure. The measures are discretized by integrating over intervals of length 1/1024. (a) One random sample of the discretized Wiener measure. The Gaussian seed have the scaling factor $1/\sqrt{1024}$. (b) One random sample of the discretized Cauchy measure. The Cauchy seed have the scaling factor $\tau_c/1024$. The factor $\tau_c \approx 0.416$ adjust the scale so that the the measure of the total region $[\,0,\,1\,]$ have the same 90'th percentile for the two measures.

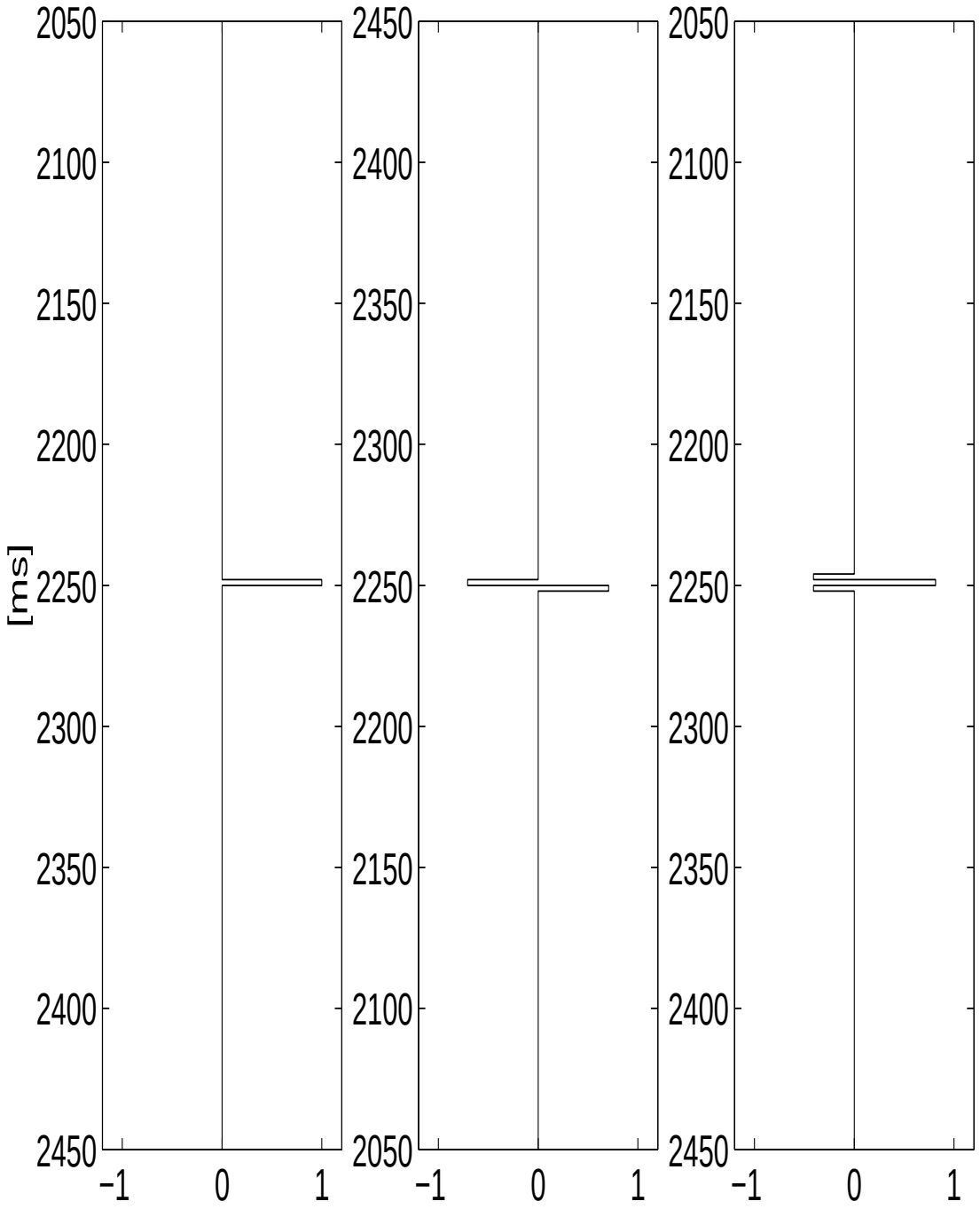

Figure 21: Directions for multi directional Gibbs sampler. The pool of basis vectors used in the multi directional Gibbs sampler, are translations of the functions above. The resolution of the figure is such that each plateau correspond to the size of one vector component.

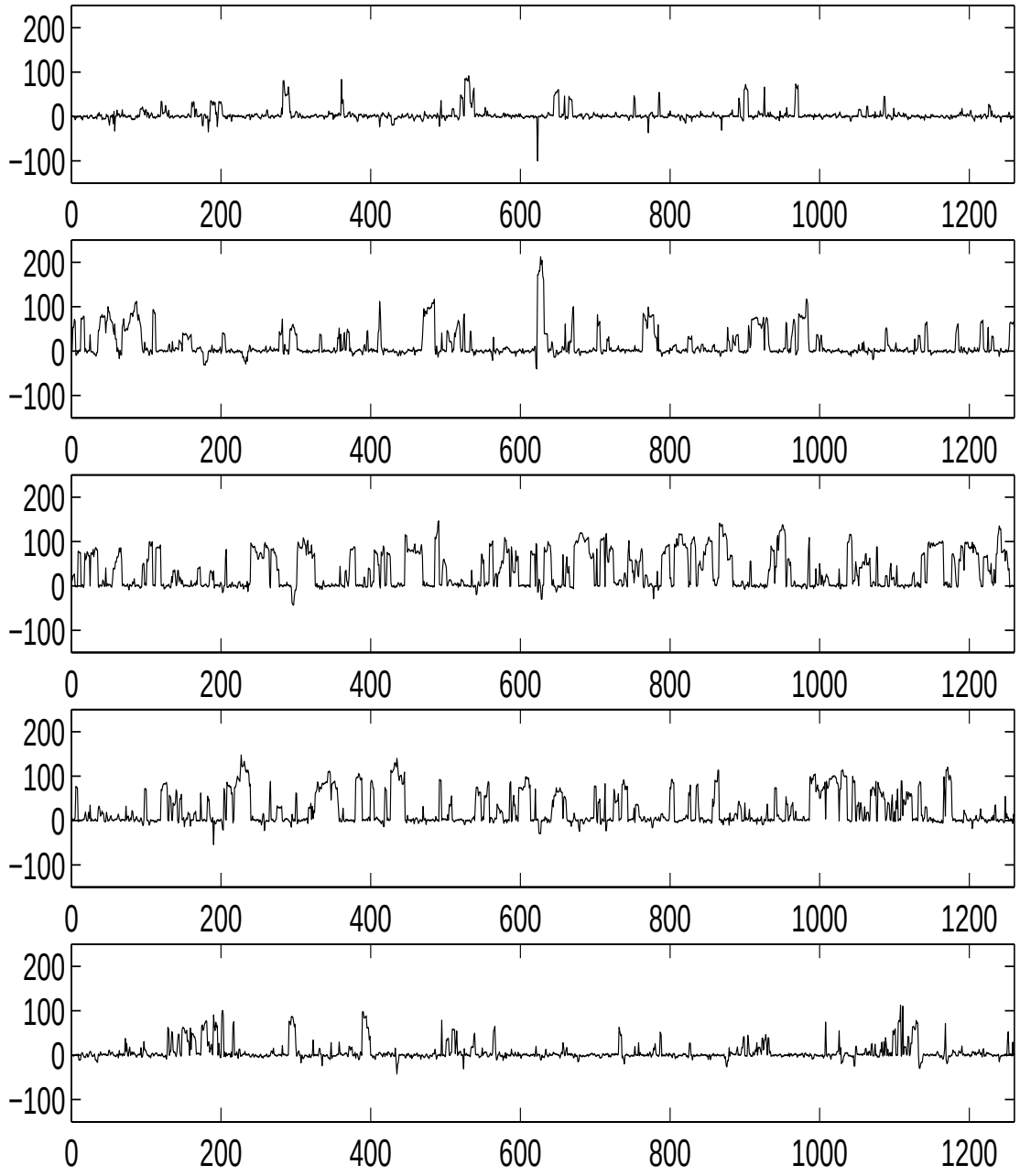

Figure 22: Mixing of multi directional Gibbs sampling, for the Sleipner Øst Field. Time traces for the the Cauchy seed. The sampled seed value that correspond to the leap value at 2380 ms is in the middle plot, the two nearest neighbors at both sides are above and below. Note that the scale of the figure is for the Cauchy seed. To get the scale that apply to the derivative of the logarithm of the material parameters, the seed should be divided by 2 due to the kernel in Figure 6(a) and multiplied by the second column in Table 1, due to the relation in Expression (4).