

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

Arild Buland^{1,2*} and Henning Omre²

¹Statoil Research Centre, Postuttak, 7005 Trondheim, and ²Norwegian University of Science and Technology, 7491 Trondheim, Norway

Received July 2002, revision accepted June 2003

ABSTRACT

The main objective of the AVO inversion is to obtain posterior distributions for P-wave velocity, S-wave velocity and density from specified prior distributions, seismic data and well-log data. The inversion problem also involves estimation of a seismic wavelet and the seismic-noise level. The noise model is represented by a zero mean Gaussian distribution specified by a covariance matrix. A method for joint AVO inversion, wavelet estimation and estimation of the noise level is developed in a Bayesian framework. The stochastic model includes uncertainty of both the elastic parameters, the wavelet, and the seismic and well-log data. The posterior distribution is explored by Markov-chain Monte-Carlo simulation using the Gibbs' sampler algorithm. The inversion algorithm has been tested on a seismic line from the Heidrun Field with two wells located on the line. The use of a coloured seismic-noise model resulted in about 10% lower uncertainties for the P-wave velocity, S-wave velocity and density compared with a white-noise model. The uncertainty of the estimated wavelet is low. In the Heidrun example, the effect of including uncertainty of the wavelet and the noise level was marginal with respect to the AVO inversion results.

INTRODUCTION

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

by the posterior distribution, which provides both the most probable solution and information about the corresponding uncertainty.

Amplitude versus offset (AVO) inversion can be used to extract information about the elastic subsurface parameters utilizing the angle dependency of the reflection coefficient (Smith and Gidlow 1987; Hampson and Russell 1990; Buland *et al.* 1996; Gouveia and Scales 1998; Wang 1999). The AVO inversion problem can be linearized if an appropriate processing sequence is applied to the seismic data prior to the inversion. Important elements in such processing are the removal of the moveout, multiples, and the effect of geometrical spreading and absorption. The seismic data should be prestack migrated (Wang, White and Pratt 2000; Buland and Landrø 2001) and transformed from offset to reflection angle. A Bayesian linearized AVO inversion is defined in Buland and Omre (2003a), where an explicit analytical form of the posterior distribution is derived under Gaussian model assumptions. The explicit analytical form of the posterior distribution provides a computationally fast inversion method suitable for inversion of large

*E-mail: abu@statoil.com

3D seismic data sets. However, the elastic parameters are not laterally coupled, so the inversion is performed independently for each bin gather. Furthermore, the seismic wavelet and the noise covariance are assumed to be known prior to the inversion. A Bayesian method for estimation of the wavelet and the noise covariance is presented in Buland and Omre (2003b).

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

METHODOLOGY

An isotropic, elastic medium is described completely by the material parameters $\{\alpha(\mathbf{x}, t), \beta(\mathbf{x}, t), \rho(\mathbf{x}, t)\}$, where α, β and ρ are P-wave velocity, S-wave velocity and density, \mathbf{x} is the lateral location, and t is the two-way vertical seismic traveltime. A weak contrast reflectivity function for PP reflections is (Aki and Richards 1980; Buland and Omre 2003a)

$$c(\mathbf{x}, t, \theta) = a_\alpha(\mathbf{x}, t, \theta) \frac{\partial}{\partial t} \ln \alpha(\mathbf{x}, t) + a_\beta(\mathbf{x}, t, \theta) \frac{\partial}{\partial t} \ln \beta(\mathbf{x}, t) + a_\rho(\mathbf{x}, t, \theta) \frac{\partial}{\partial t} \ln \rho(\mathbf{x}, t), \tag{1}$$

where θ is the angle of reflection, $a_\alpha = (1 + \tan^2 \theta)/2$, $a_\beta = -4(\beta/\alpha)^2 \sin^2 \theta$ and $a_\rho = (1 - 4(\beta/\alpha)^2 \sin^2 \theta)/2$. The inversion algorithm requires that a_α, a_β and a_ρ are defined from a

prior known background model. Motivated by the form of the reflectivity function in expression (1), let the unknown model parameter vector be

$$\mathbf{m}(\mathbf{x}, t) = [\ln \alpha(\mathbf{x}, t), \ln \beta(\mathbf{x}, t), \ln \rho(\mathbf{x}, t)]^T, \tag{2}$$

where superscript T denotes transpose.

The seismic data are represented by the convolutional model

$$d_{\text{obs}}(\mathbf{x}, t, \theta) = \int s(\tau, \theta) c(\mathbf{x}, t - \tau, \theta) d\tau + e_d(\mathbf{x}, t, \theta), \tag{3}$$

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

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

The stochastic dependency model

The variables and the problem structure are graphically displayed by a directed acyclic graph (DAG) in Fig. 1 (see e.g.

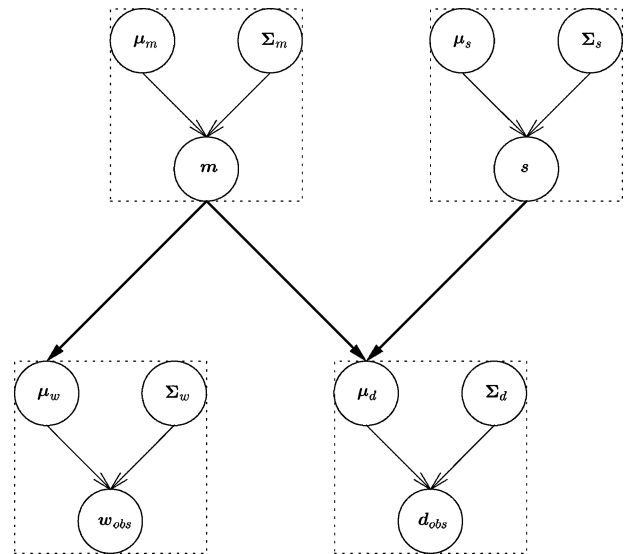


Figure 1 The stochastic model represented by a directed acyclic graph. The nodes represent stochastic variables, the thin arrows represent probability dependencies, and the thick arrows represent deterministic relationships.

Lauritzen 1996; Spiegelhalter *et al.* 1996). All nodes in the graph represent stochastic variables, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote expectation vectors and covariance matrices. The arrows represent the causal dependency structure, where thin lines indicate probability dependencies and thick lines represent deterministic relationships. The DAG can be interpreted as a family tree where the parent nodes point to its children. When probabilistic dependencies are considered, nodes connected by a deterministic link merge into a single node. The expectation vectors $\boldsymbol{\mu}_w$ and $\boldsymbol{\mu}_d$ are deterministically determined from \mathbf{m} and from \mathbf{m} and \mathbf{s} , respectively. The parents of the well-log data are therefore $\text{Pa}(\mathbf{w}_{\text{obs}}) = \{\mathbf{m}, \boldsymbol{\Sigma}_w\}$, and the parents of the seismic data are $\text{Pa}(\mathbf{d}_{\text{obs}}) = \{\mathbf{m}, \mathbf{s}, \boldsymbol{\Sigma}_d\}$, where the notation $\text{Pa}(v)$ denotes the parents of a node v . Before any data are observed, the marginal distributions of the parents are independent, but when their child is observed, the properties of the parents become dependent. An example is \mathbf{m} and \mathbf{s} which are *a priori* independent, but become dependent when the observed seismic data \mathbf{d}_{obs} are included in the model.

The likelihood model

The error term in the convolutional model in expression (3) is assumed to be zero mean Gaussian with covariance $\boldsymbol{\Sigma}_d$. The likelihood model for the seismic data \mathbf{d}_{obs} is then Gaussian, compactly denoted

$$\mathbf{d}_{\text{obs}} | \boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d \sim \mathcal{N}_{n_d}(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d), \quad (4)$$

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

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

$$\mathbf{w}_{\text{obs}} = \mathbf{P}\mathbf{m} + \mathbf{e}_w, \quad (5)$$

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

$$\mathbf{w}_{\text{obs}} | \boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w \sim \mathcal{N}_{n_w}(\boldsymbol{\mu}_w, \boldsymbol{\Sigma}_w), \quad (6)$$

where n_w is the dimension of \mathbf{w}_{obs} and $\boldsymbol{\mu}_w = \mathbf{P}\mathbf{m}$. The well locations relative to the seismic data are assumed to be correct. Also the conversion of the well logs from depth to seismic traveltimes is assumed to be correct. In principle, this uncertainty

related to the depth-to-time transform of the well logs could have been included in the present method by simulation of possible shift, stretch and squeeze of the log time axis (Buland and Omre 2003b).

The prior model

The model parameters \mathbf{m} are *a priori* assumed to be Gaussian,

$$\mathbf{m} | \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m \sim \mathcal{N}_{n_m}(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m), \quad (7)$$

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

Also the wavelet \mathbf{s} is *a priori* modelled with a Gaussian distribution,

$$\mathbf{s} | \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s \sim \mathcal{N}_{n_s}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s), \quad (8)$$

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

The DAG in Fig. 1 represents a hierarchical Bayesian model, where the expectations and covariances are stochastic. The prior expressions for \mathbf{m} and \mathbf{s} in expressions (7) and (8) can be regarded as the first level of a hierarchical prior model, the structural portion, while the prior distributions for the expectation vectors and covariance matrices form the second level, the subjective portion of the prior (see Robert 1994; Carlin 1996).

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

$$\boldsymbol{\mu}_m \sim \mathcal{N}_{n_m}(\boldsymbol{\mu}_{\boldsymbol{\mu}_m}, \boldsymbol{\Sigma}_{\boldsymbol{\mu}_m}), \quad (9)$$

$$\boldsymbol{\mu}_s \sim \mathcal{N}_{n_s}(\boldsymbol{\mu}_{\mu_s}, \boldsymbol{\Sigma}_{\mu_s}), \quad (10)$$

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

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

$$\boldsymbol{\Sigma}_i = \sigma_i^2 \boldsymbol{\Sigma}_{0,i}, \quad (11)$$

where i is one of $\{m, s, w, d\}$. The uncertainty of the unknown variance factor σ_i^2 is modelled by the inverse gamma distribution,

$$\sigma_i^2 \sim \mathcal{IG}(\gamma_i, \lambda_i), \quad (12)$$

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

The posterior model

The objective of the inversion is to obtain the posterior distributions for the variables involved, based on the observed seismic data \mathbf{d}_{obs} and the well logs \mathbf{w}_{obs} . Let the vector \mathbf{v} represent the unknown quantities, that is, all variables in the DAG in Fig. 1 except \mathbf{d}_{obs} and \mathbf{w}_{obs} . Further, let \mathbf{o} be a vector containing the observed data,

$$\mathbf{o} = \begin{bmatrix} \mathbf{w}_{\text{obs}} \\ \mathbf{d}_{\text{obs}} \end{bmatrix}. \quad (13)$$

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

$$p(\mathbf{v} | \mathbf{o}) = \frac{p(\mathbf{v}, \mathbf{o})}{p(\mathbf{o})}, \quad (14)$$

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

$$p(\mathbf{v}, \mathbf{o}) = \prod_{x_i \in \{\mathbf{v}, \mathbf{o}\}} p(x_i | \text{Pa}(x_i)). \quad (15)$$

An expression for the posterior distribution is now defined from expressions (14) and (15), and the DAG in Fig. 1,

$$\begin{aligned} & p(\mathbf{m}, \mathbf{s}, \boldsymbol{\mu}_m, \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_m, \boldsymbol{\Sigma}_s, \boldsymbol{\Sigma}_w, \boldsymbol{\Sigma}_d | \mathbf{d}_{\text{obs}}, \mathbf{w}_{\text{obs}}) \\ & \propto p(\mathbf{d}_{\text{obs}} | \mathbf{m}, \mathbf{s}, \boldsymbol{\Sigma}_d) p(\mathbf{w}_{\text{obs}} | \mathbf{m}, \boldsymbol{\Sigma}_w) p(\mathbf{m} | \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \\ & p(\mathbf{s} | \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s) p(\boldsymbol{\mu}_m) p(\boldsymbol{\mu}_s) p(\boldsymbol{\Sigma}_m) p(\boldsymbol{\Sigma}_s) p(\boldsymbol{\Sigma}_w) p(\boldsymbol{\Sigma}_d), \end{aligned} \quad (16)$$

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

The Gibbs' sampler algorithm

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

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

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

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

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

$$v_1^{(i)} \sim p(v_1 | \mathbf{v}_{-1}, \mathbf{o}),$$

$$\vdots$$

$$v_j^{(i)} \sim p(v_j | \mathbf{v}_{-j}, \mathbf{o}),$$

$$\vdots$$

$$v_n^{(i)} \sim p(v_n | \mathbf{v}_{-n}, \mathbf{o}),$$

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

INVERSION EXAMPLE OF HEIDRUN DATA

The Heidrun Field is an oil and gas field, offshore mid-Norway. Geologically, the field comprises a heavily faulted

and eroded horst block of early to mid-Jurassic clastic deposits. The reservoirs are located in the shallow, clean Fangst Group sandstones, and in the deeper and more heterogeneous sandstones of the Tilje and Åre Formations.

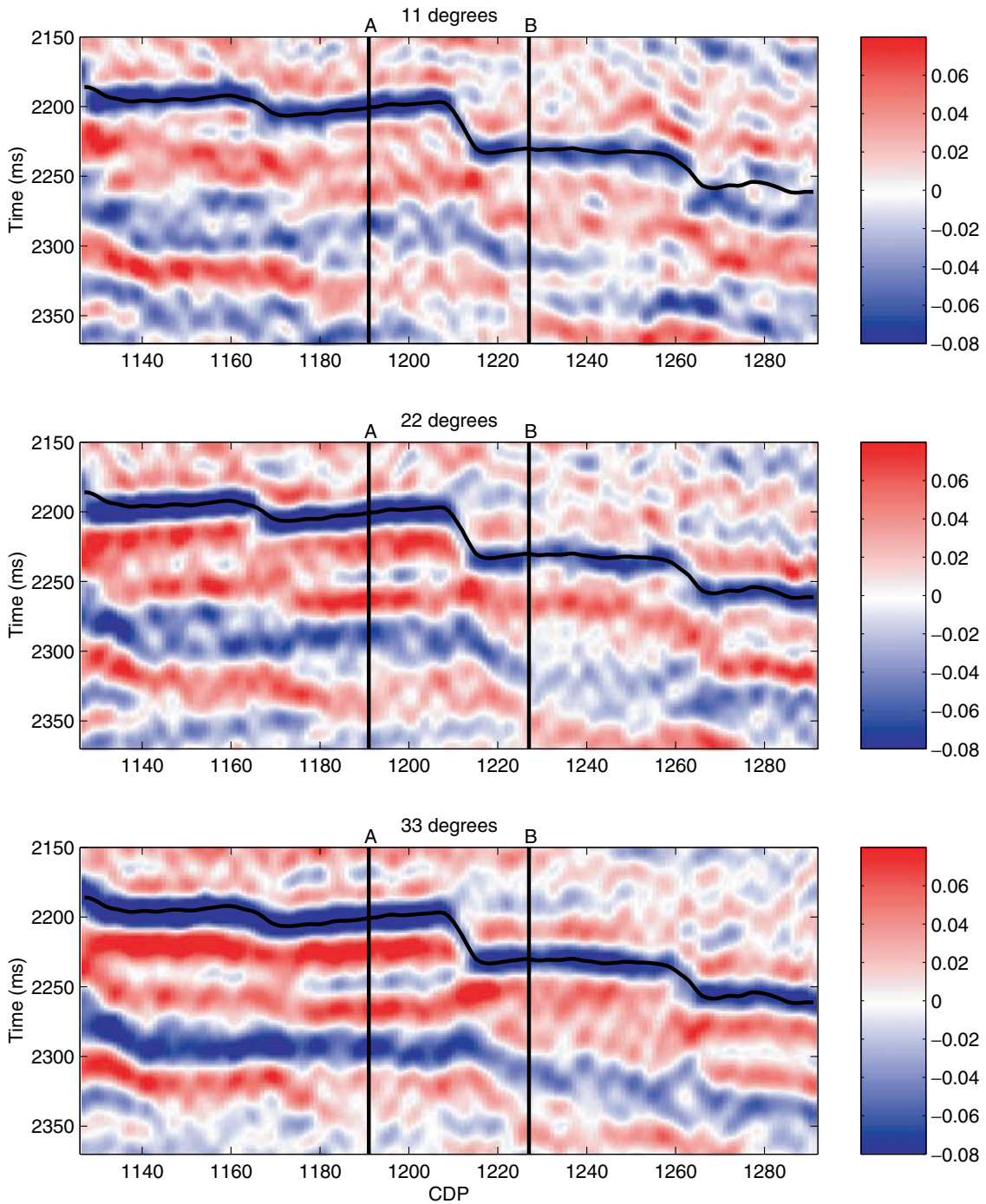


Figure 2 The seismic data represented by angle stacks: 11° (top), 22° (middle) and 33° (bottom).

A seismic line passing through two wells, A and B, is used in this inversion example. The seismic data consist of $n_\theta = 3$ angle stacks with average angles 11° , 22° and 33° (Fig. 2). Well A is located at CDP 1191 and well B at CDP 1227. An

interpretation of top reservoir (Fangst Group) is shown by a black line. A 300 ms time window is inverted, where the top and bottom are parallel to the interpreted top reservoir horizon. Typical seismic peak amplitudes are of the order of 0.1.

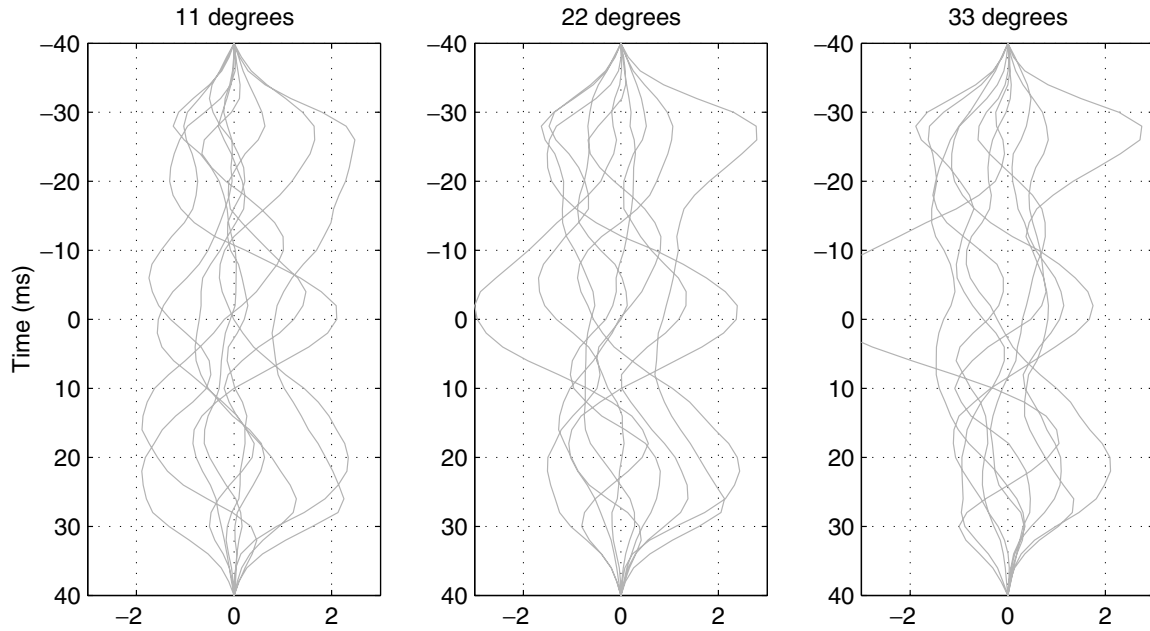


Figure 3 Wavelets simulated from the prior distribution.

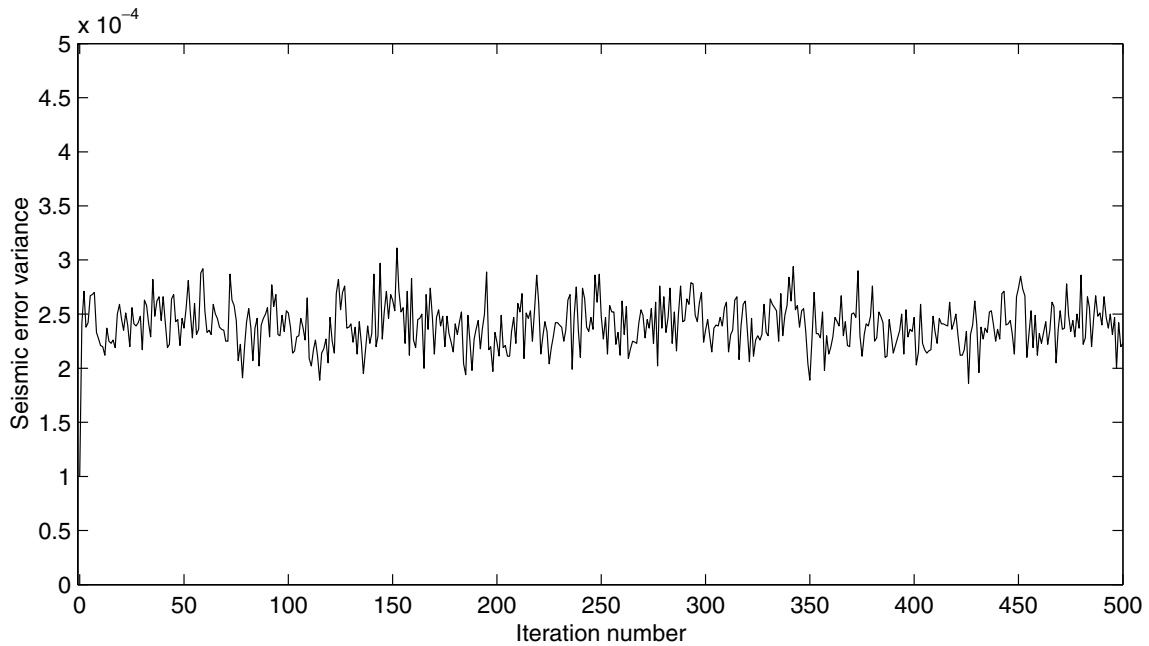


Figure 4 Monitor plot of the seismic white-noise variance σ_d^2 .

The prior model

Among the hierarchical parameters, we consider the unknown seismic error covariance matrix Σ_d to be the most interesting.

The estimation of both the wavelet s and the model parameter vector m with corresponding uncertainties depends on Σ_d . Unfortunately, the estimation of Σ_d is usually not a trivial problem. The dominating noise in processed seismic data is

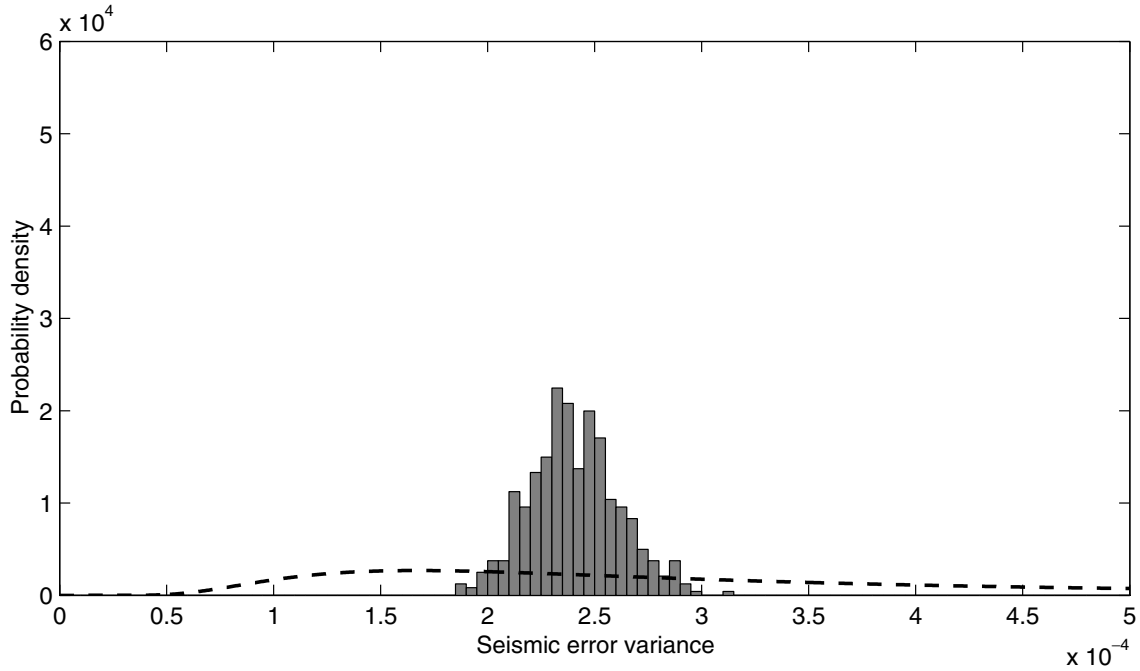


Figure 5 Posterior distribution for the seismic white-noise variance σ_d^2 . The *a priori* distribution is shown dotted.

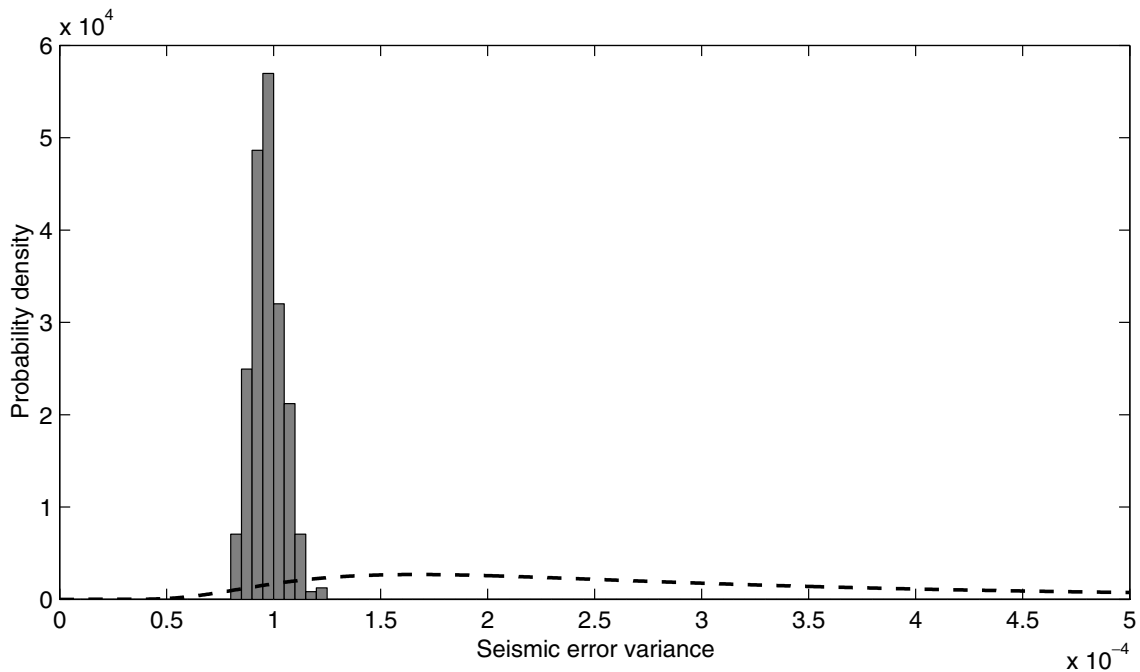


Figure 6 Posterior distribution for the seismic error variance σ_d^2 for coloured noise. The *a priori* distribution is shown dotted.

usually source-generated noise, for example remaining multiples or processing artefacts. Such noise components usually have a smooth waveform similar to the waveform of the primary events. However, if well logs are available, the misfit between the synthetic seismic well-log response and the real seismic data can be used to estimate Σ_d , or a set of parameters which determines the complete Σ_d . Three different

seismic-noise models are defined below, numbered from 1 to 3. The covariance matrices are here defined for a single CDP gather with $n_\theta = 3$ traces, and we assume no correlation between different CDP positions. In contrast, Σ_d in expression (4) is the covariance for the complete seismic data set. Accordingly, n_d is here reduced to be the dimension of a CDP gather.

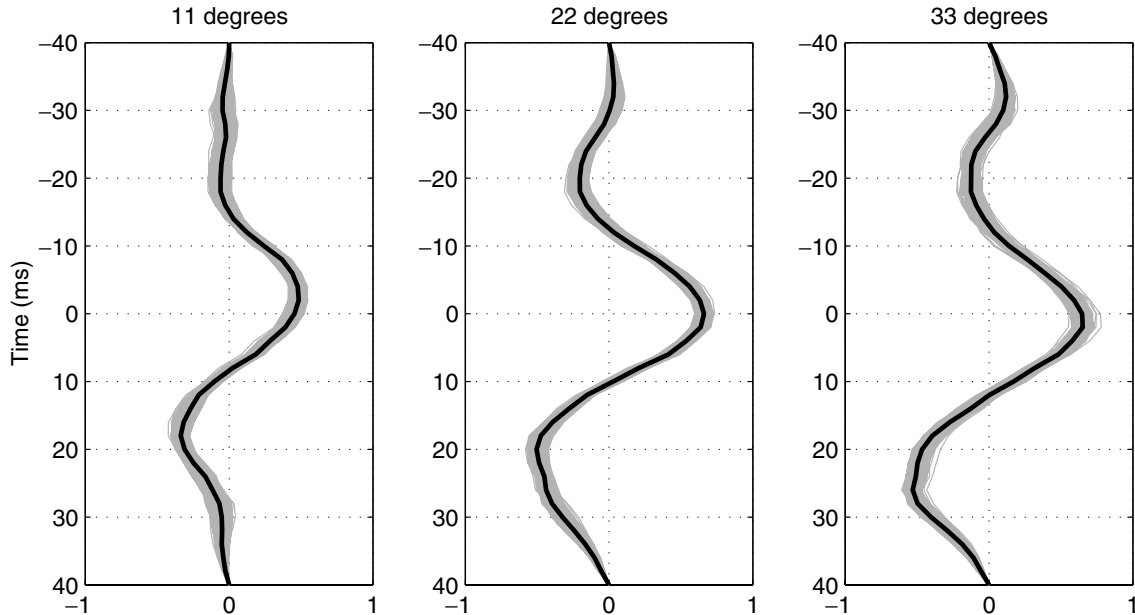


Figure 7 Simulated wavelets (grey lines) and the posterior mean wavelet (black line) for noise model 1.

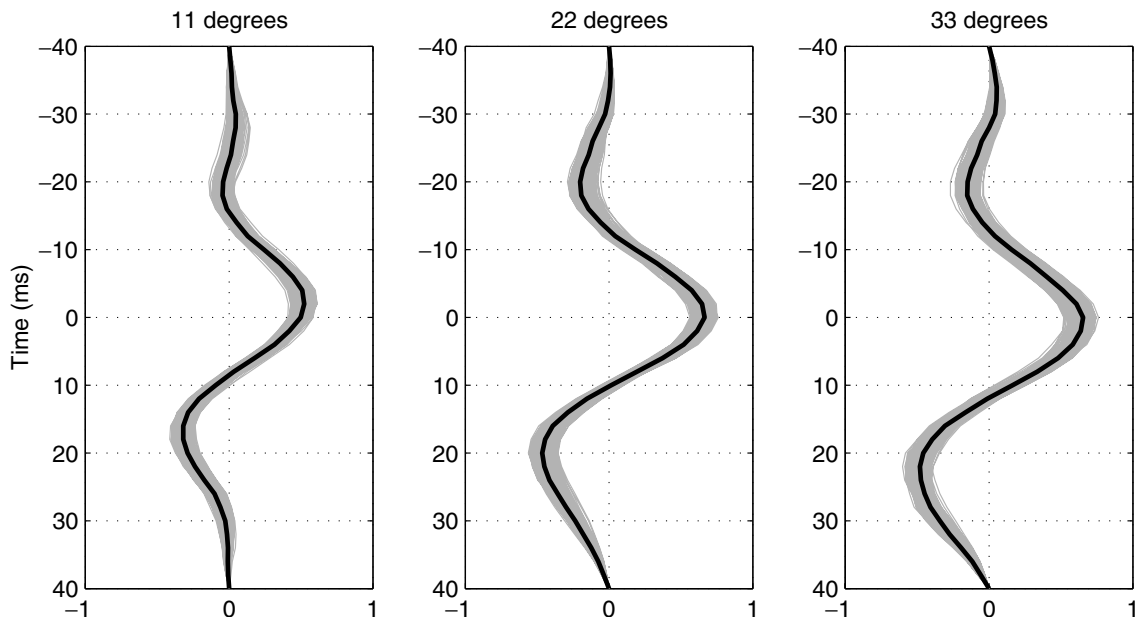


Figure 8 Simulated wavelets (grey lines) and the posterior mean wavelet (black line) for noise model 2.

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

$$\Sigma_{d,1} = \sigma_d^2 \mathbf{I}_{n_d}, \tag{17}$$

where \mathbf{I}_{n_d} is an $n_d \times n_d$ identity matrix. Our *a priori* knowledge about the seismic-noise level is vague. The *a priori* uncertainty is described by the inverse gamma distribution

$$\sigma_d^2 \sim \mathcal{IG}(2, 0.0005), \tag{18}$$

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

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

$$\Sigma_{d,2} = \sigma_d^2 \mathbf{I}_{n_\theta} \otimes \Upsilon_d, \tag{19}$$

where each of the elements in the 3×3 identity matrix \mathbf{I}_{n_θ} is multiplied with the temporal $n_t \times n_t$ correlation matrix Υ_d , with $n_t = 151$ being the number of time samples. This means that we assume that the error variance is equal for the three angle stacks, and that there is no correlation between them. The unknown variance σ_d^2 is modelled by the inverse gamma

distribution as in expression (18). Note that noise model 1 can be written in a similar form, $\Sigma_{d,1} = \sigma_d^2 \mathbf{I}_{n_\theta} \otimes \mathbf{I}_{n_t}$.

Noise model 3 has coloured covariance with a separate unknown variance factor for each angle stack and unknown correlation coefficients between the angles. The covariance matrix is obtained by substituting the diagonal matrix $\sigma_d^2 \mathbf{I}_{n_\theta}$ in expression (19) with an unknown $n_\theta \times n_\theta$ covariance matrix Σ_θ , such that

$$\Sigma_{d,3} = \Sigma_\theta \otimes \Upsilon_d. \tag{20}$$

The covariance matrix Σ_θ is *a priori* assigned an inverted Wishart distribution with 5 degrees of freedom,

$$\Sigma_\theta \sim \mathcal{IW}_{n_\theta}(0.0001\mathbf{I}_{n_\theta}, 5) \tag{21}$$

(see Appendix C). With only 5 degrees of freedom, the prior distribution will only have a marginal effect on the conditional distribution for Σ_θ . Here, the relative influence of the prior distribution versus the sample covariance is about 1 to 60. If the prior knowledge about Σ_θ is more precise, the relative influence of the prior distribution can be increased by increasing the degree of freedom.

The wavelet is *a priori* assumed to be Gaussian with unknown expectation and covariance. The wavelet length is set at 80 ms, which is sufficient for a main centre loop and two side loops. A separate wavelet is estimated for each of the three angle stacks, such that $n_s = 3 \times 41$ with 2 ms sampling. The expected temporal smoothness is imposed through a

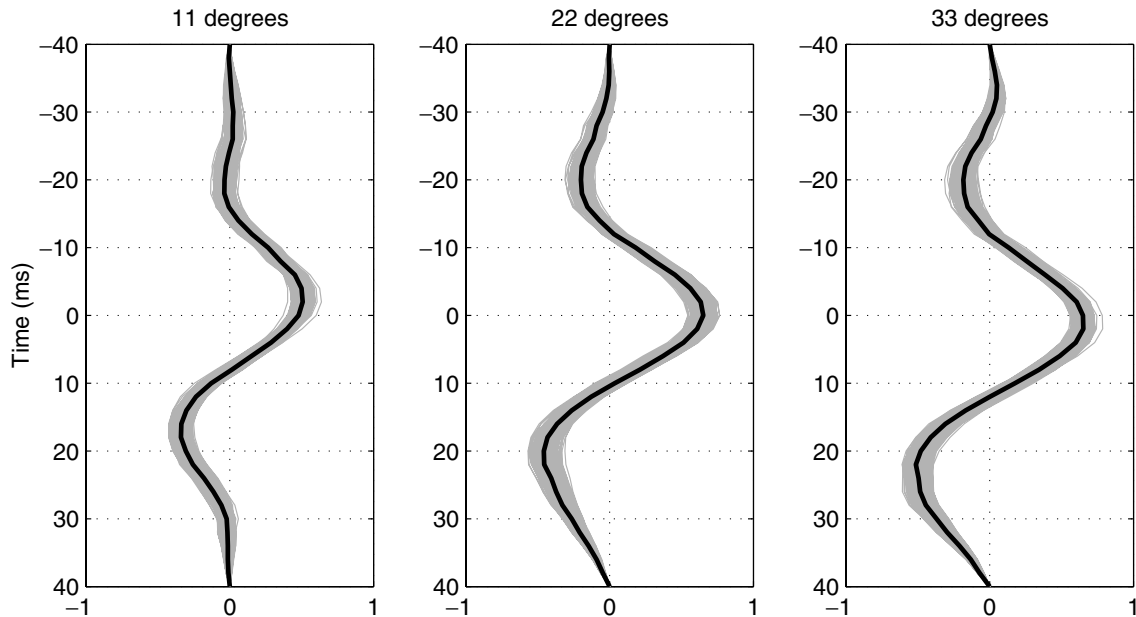


Figure 9 Simulated wavelets (grey lines) and the posterior mean wavelet (black line) for noise model 3.

covariance matrix $\Sigma_{0,s}$ by a second-order exponential correlation function with range 14 ms. Note that this correlation function is identical to the correlation used to model the smoothness of the coloured seismic noise. Furthermore, it is

imposed so that the wavelets smoothly approach zero towards the first and last wavelet samples. The wavelets are also assumed to vary smoothly as a function of angle of reflection. This is imposed through $\Sigma_{0,s}$ by a second-order exponential

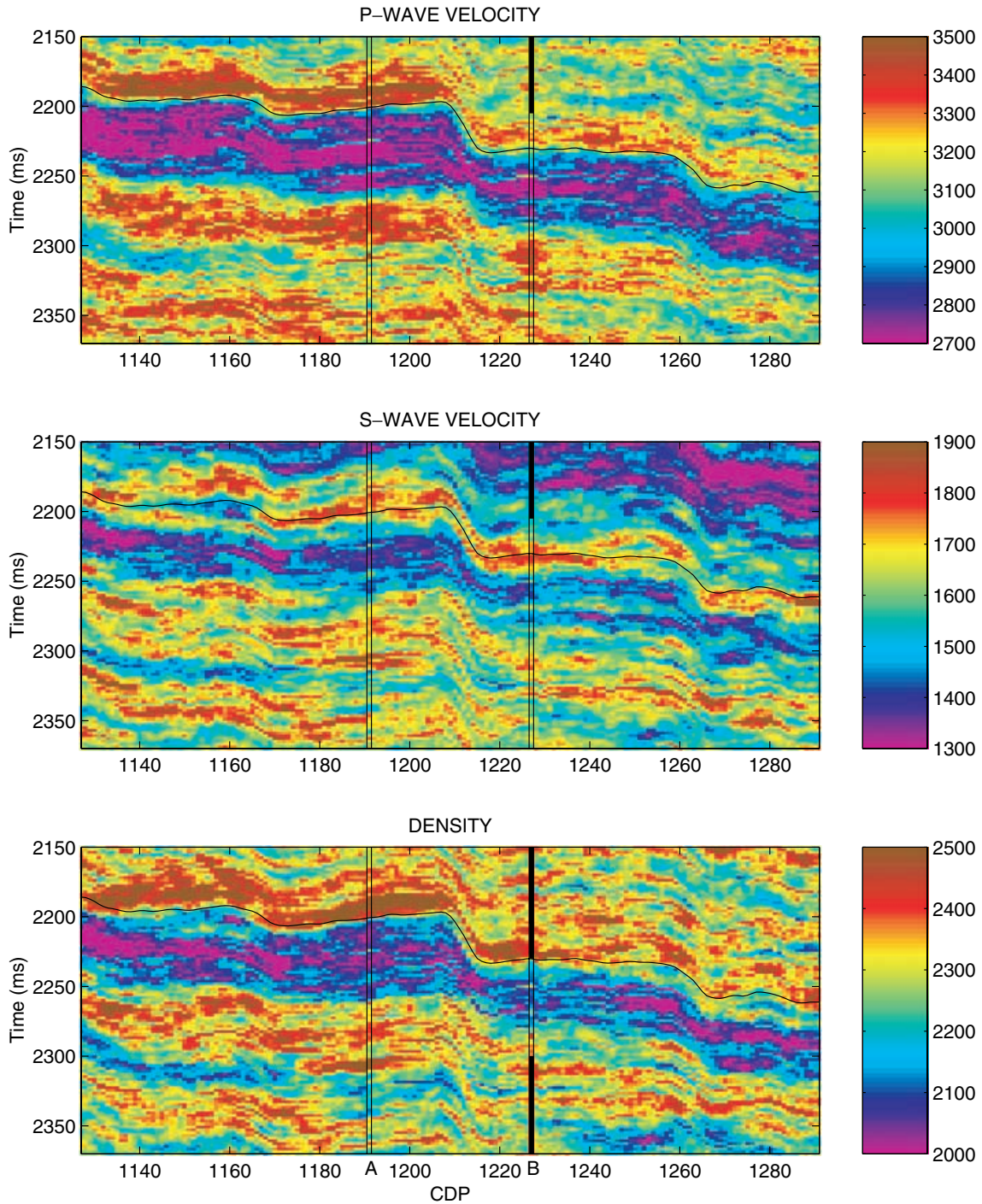


Figure 10 Simulated solution conditioned to seismic and well-log data.

correlation function with range 30° . The prior distribution for the wavelet expectation is

$$\boldsymbol{\mu}_s \sim \mathcal{N}_{n_s}(\mathbf{0}, \boldsymbol{\Sigma}_{\mu_s}), \quad (22)$$

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

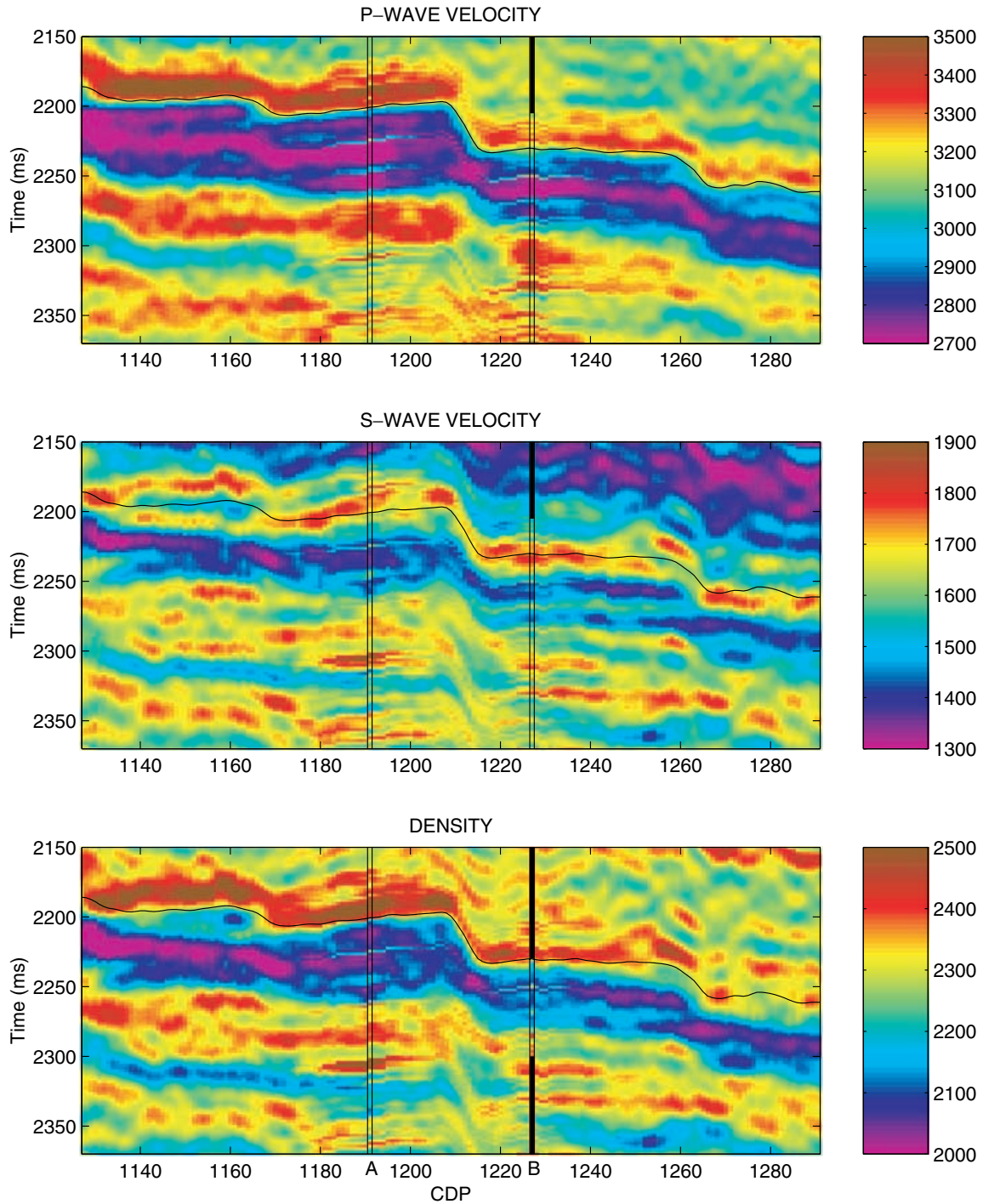


Figure 11 Posterior mean solution conditioned to seismic and well-log data.

covariance Σ_s is

$$\Sigma_s = \sigma_s^2 \Sigma_{0,s}, \quad (23)$$

where the *a priori* model for the unknown variance is

$$\sigma_s^2 \sim \mathcal{IG}(2, 0.001), \quad (24)$$

with expectation 0.001 and undefined (infinite) variance. To illustrate the defined prior model, a set of wavelets simulated from the prior distribution is shown in Fig. 3. Except for the imposed smoothness and the amplitude decay towards -40 ms and 40 ms, the prior model is flexible with respect to the shape of the wavelets.

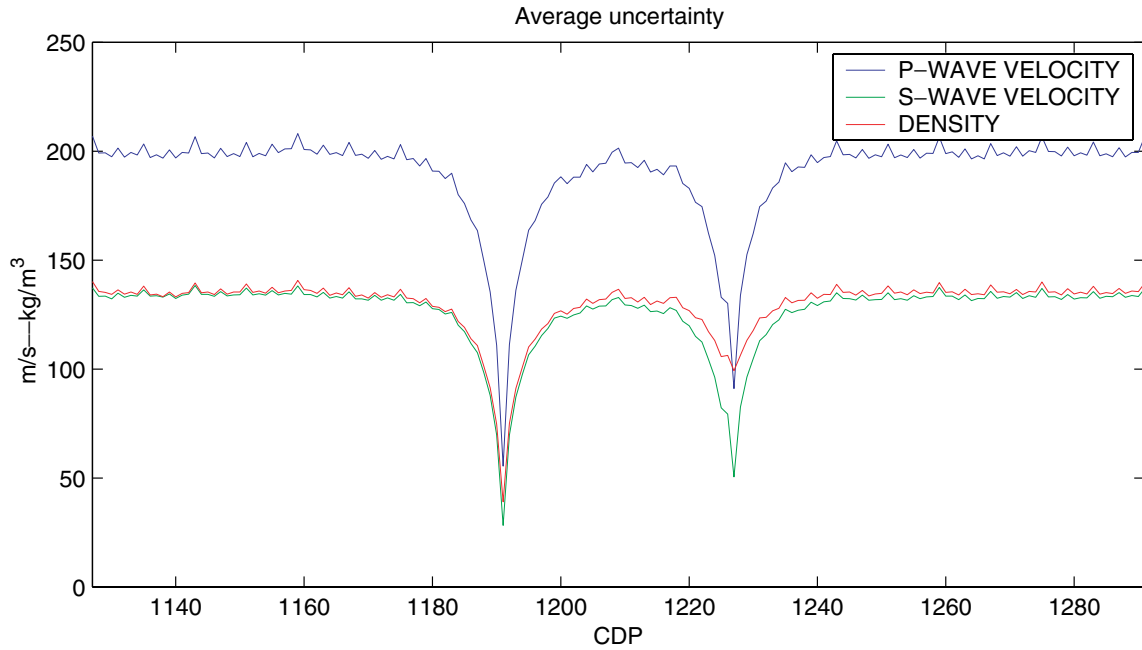


Figure 12 Average posterior uncertainty for P-wave velocity (blue), S-wave velocity (green) and density (red) for noise model 1.

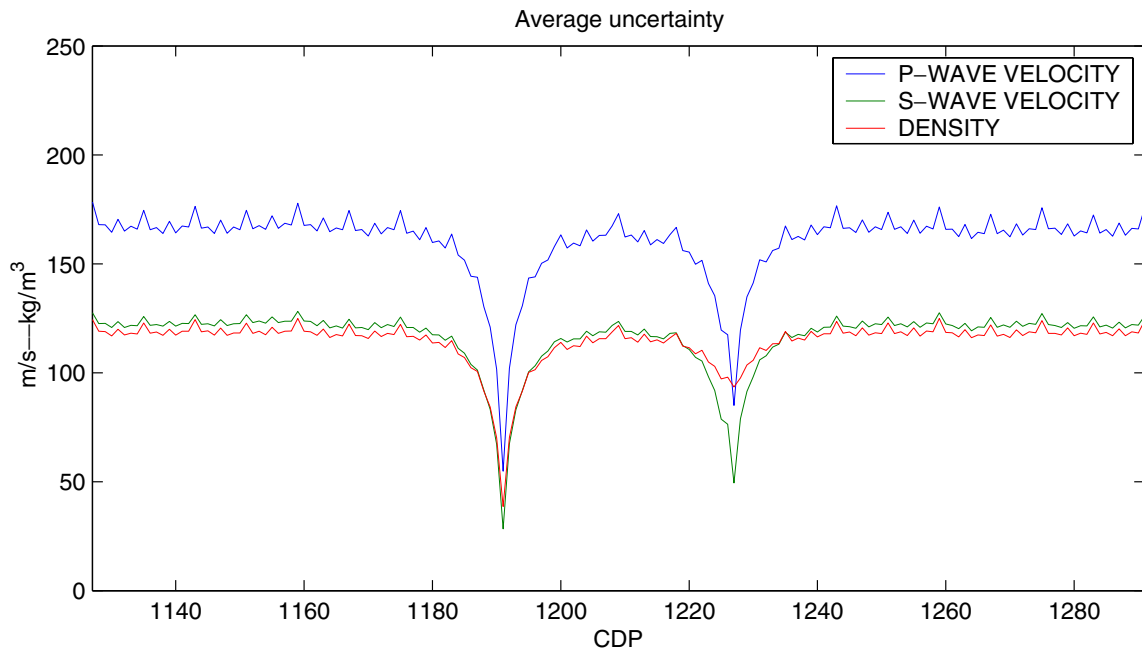


Figure 13 Average posterior uncertainty for P-wave velocity (blue), S-wave velocity (green) and density (red) for noise model 2.

The prior model for \mathbf{m} is modelled by a Gaussian distribution specified by the expectation $\boldsymbol{\mu}_m$ and the covariance $\boldsymbol{\Sigma}_m$. A thorough presentation of the prior model for \mathbf{m} can be found in Buland and Omre (2003a). In this example, the prior model for \mathbf{m} is determined from the well logs and then fixed. The expectation $\boldsymbol{\mu}_m$ can be considered to be a slowly varying background model for $\ln \alpha(\mathbf{x}, t)$, $\ln \beta(\mathbf{x}, t)$ and $\ln \rho(\mathbf{x}, t)$. The temporal variation of $\boldsymbol{\mu}_m$ should be smooth, but of sufficiently high frequency to cover the missing low frequencies in the band-limited seismic data. In this example, the prior expectation is vertically slowly varying but laterally constant, which means constant along structures parallel to the interpreted top reservoir horizon. The covariance function is stationary and homogeneous, and can be factorized as

$$\boldsymbol{\Sigma}_m(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2) = \boldsymbol{\Sigma}_{0,m} v_m(\xi, \tau), \quad (25)$$

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

$$\boldsymbol{\Sigma}_{0,m} = \begin{bmatrix} \sigma_\alpha^2 & \sigma_\alpha \sigma_\beta v_{\alpha\beta} & \sigma_\alpha \sigma_\rho v_{\alpha\rho} \\ \sigma_\alpha \sigma_\beta v_{\alpha\beta} & \sigma_\beta^2 & \sigma_\beta \sigma_\rho v_{\beta\rho} \\ \sigma_\alpha \sigma_\rho v_{\alpha\rho} & \sigma_\beta \sigma_\rho v_{\beta\rho} & \sigma_\rho^2 \end{bmatrix}. \quad (26)$$

The diagonal elements of $\boldsymbol{\Sigma}_{0,m}$ are the variances, and $v_{\alpha\beta}$, $v_{\alpha\rho}$ and $v_{\beta\rho}$ are the correlations between $\ln \alpha(\mathbf{x}, t)$, $\ln \beta(\mathbf{x}, t)$ and $\ln \rho(\mathbf{x}, t)$, respectively. From the well logs, estimates of the elements in $\boldsymbol{\Sigma}_{0,m}$ are obtained by standard estimators. The estimated variances for $\ln \alpha$, $\ln \beta$ and $\ln \rho$ are $\widehat{\sigma}_\alpha^2 = 0.0026$,

$\widehat{\sigma}_\beta^2 = 0.0034$ and $\widehat{\sigma}_\rho^2 = 0.0020$, and the estimated correlation coefficients are $\widehat{v}_{\alpha\beta} = 0.76$, $\widehat{v}_{\alpha\rho} = 0.77$ and $\widehat{v}_{\beta\rho} = 0.81$. The vertical variability is modelled by a sum of an exponential second-order correlation function with range 2 ms and a normalized second derivative of an exponential second-order correlation function with range 10 ms (Buland and Omre 2003a). The lateral coupling of the model parameters is imposed by a first-order exponential correlation function with range 125 m, i.e. 10 CDP positions.

In the well positions, the well logs represent measurements of the model parameters \mathbf{m} . The covariance of the possible well-log errors is set at

$$\boldsymbol{\Sigma}_w = \sigma_w^2 \mathbf{I}, \quad (27)$$

with $\sigma_w^2 = 0.0001$, such that the variance is fixed and identical for the $\ln \alpha$, $\ln \beta$ and $\ln \rho$ logs. A non-informative prior distribution should not be specified for σ_w^2 since the seismic-error term in the convolutional model is not completely separable from errors corresponding to the well-log errors. The prior model for σ_w^2 should therefore be specified exactly, or at least relatively precisely (Buland and Omre 2003b).

Simulation results

The posterior distribution is explored by 500 iterations with the Gibbs' sampler algorithm. MCMC algorithms generally

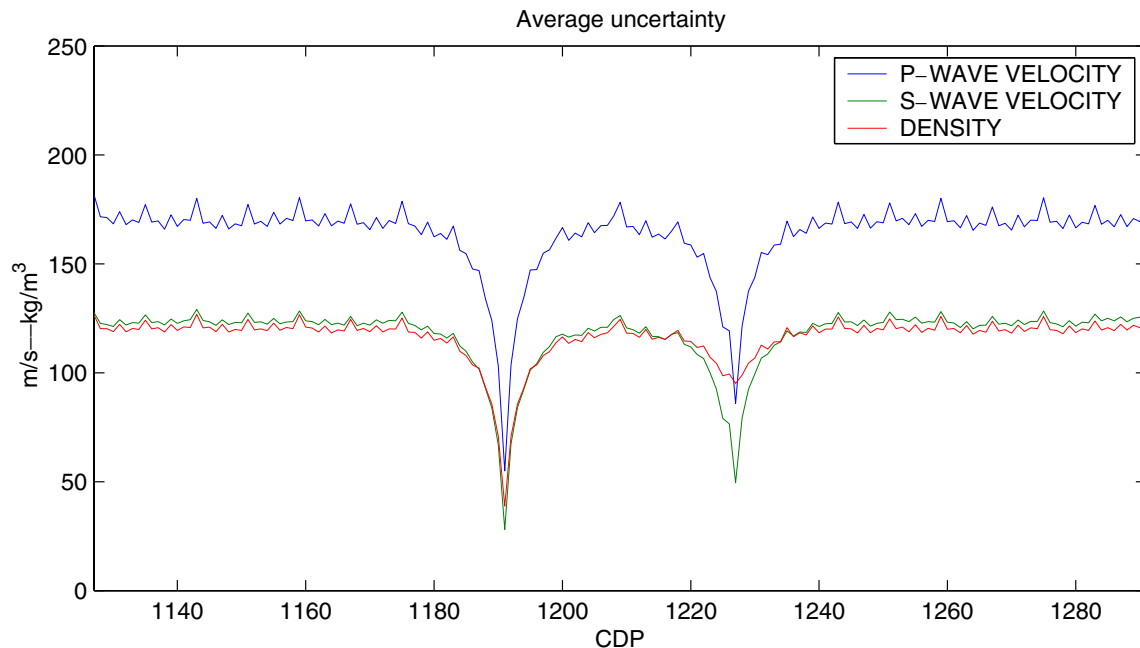


Figure 14 Average posterior uncertainty for P-wave velocity (blue), S-wave velocity (green) and density (red) for noise model 3.

need some initial iterations before convergence to the posterior distribution is reached. The convergence can be evaluated by monitoring the involved variables using different initial values. In this problem, the convergence is fast; only some few

burn-in iterations are needed. To be certain, the first 20 iterations are considered to be burn-in iterations, and these are not used in the calculations. For noise model 1 in expression (17), the simulated seismic error variance is shown as a function

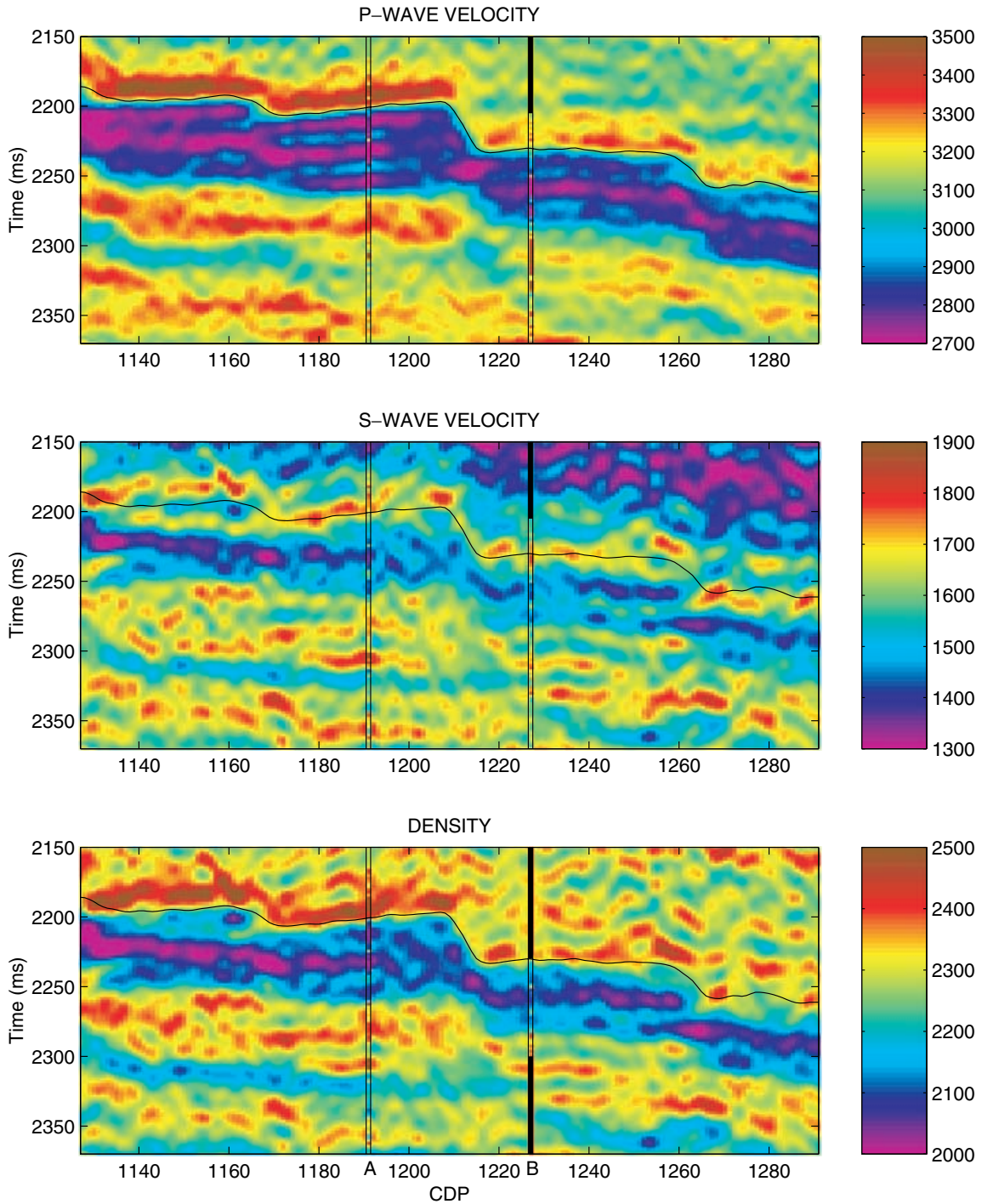


Figure 15 Posterior mean solution conditioned to seismic data.

of iteration number in Fig. 4. The initial variance was 0.0001, but the convergence to the level between 0.0002 and 0.0003 is reached after the first iteration. Tests with other initial values also show good convergence properties. The posterior distribution of σ_d^2 is represented by a histogram of the simulated variance values in Fig. 5. The posterior mean of the seismic white-noise variance is estimated from the MCMC samples to be 0.00024. For noise model 2 in expression (19), the posterior distribution for the variance factor is shown in Fig. 6, with a posterior mean of about 0.0001. A lower variance factor is sufficient to represent the seismic error with the coloured covariance compared with the white covariance. For noise model 3 in expression (20), the posterior mean for Σ_θ is estimated to be

$$\Sigma_\theta = 10^{-4} \begin{bmatrix} 0.91 & 0.14 & -0.21 \\ 0.14 & 1.03 & 0.56 \\ -0.21 & 0.56 & 0.99 \end{bmatrix}. \quad (28)$$

The variance factors on the diagonal of Σ_θ are about 0.0001 for the three angle stacks, as for noise model 2. The correlations are 0.15 between the error traces for 11° and 22°, 0.5 between 22° and 33°, and -0.2 between 11° and 33°. The posterior uncertainty of Σ_θ represented by two standard deviations is

$$10^{-4} \begin{bmatrix} \pm 0.21 & \pm 0.16 & \pm 0.15 \\ \pm 0.16 & \pm 0.24 & \pm 0.21 \\ \pm 0.15 & \pm 0.21 & \pm 0.24 \end{bmatrix}. \quad (29)$$

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

The simulated wavelets for the three different seismic-noise models in expressions (17), (19) and (20) are plotted in Figs 7, 8 and 9 with grey lines. The grey clouds illustrate the wavelet uncertainty. The estimated posterior mean wavelets are shown with black lines. The differences between the wavelets for the three seismic-noise models are marginal.

A simulated solution for the P-wave velocity, the S-wave velocity and the density is shown in Fig. 10 using seismic-noise model 2 in expression (19). The simulated solution represents one possible laterally consistent solution with high vertical variability. In some cases, for example in fluid-flow simulation, a set of simulated solutions with realistic variability is needed. Usually, however, it is more convenient to summarize the results in a best solution with the corresponding uncertainty, for example the posterior mean or the maximum posterior solution. The posterior mean solution obtained from the simulations is shown in Fig. 11. The corresponding

solutions using the two other seismic covariance models differ only marginally. Note that the solution has highest resolution near the two wells. The additional information obtained from the well logs reduces the uncertainty near the wells. The vertically averaged posterior uncertainty represented by two standard deviations is shown in Figs 12, 13 and 14 for each CDP position along the seismic line for the three different seismic-noise models. In comparison, the average *a priori* uncertainty is 316 m/s for the P-wave velocity, 181 m/s for the S-wave velocity and 200 kg/m³ for the density. The average uncertainty at well B is higher compared with the uncertainty at well A since the well-log data in B do not cover the complete inversion window. The jagged form of the curves is related to the approximate screening algorithm presented in Appendix E. However, we consider the deviations from a smooth curve to be small compared with the total uncertainty. The posterior uncertainty of the estimated elastic parameters is about 10–15% higher for the white-noise model than for the two coloured-noise models.

If the wavelet and the noise covariance are estimated prior to the inversion, and then plugged into the inversion as fixed quantities, the solution of the linearized AVO inversion problem is given in an explicit analytical form, and MCMC simulation is not required. In this paper, a stochastic model is defined which includes the uncertainty of \mathbf{s} and Σ_d . However, in the Heidrun example the effect of including the uncertainty of the wavelet and the noise covariance was marginal with respect to the estimated \mathbf{m} and its uncertainty. With the peaked posterior distribution for the noise variance in Fig. 6 and the low wavelet uncertainty in Fig. 8, this is not surprising. In cases where the wavelet and the covariance uncertainty are larger, the corresponding effect on the estimated P-wave velocity, the S-wave velocity, and the density may be more distinct.

If the lateral coupling of the model is abandoned, the AVO inversion cannot be conditioned to the well-log data, except exactly in the well positions. The posterior mean solution with no lateral coupling and with fixed wavelet and noise covariance is shown in Fig. 15, and the corresponding posterior uncertainty is shown in Fig. 16. The differences between the MCMC solution in Fig. 11 and the analytically obtained solution in Fig. 15 are minor, except near the wells. The uncertainty level is close to the maximum uncertainty in Fig. 13. The difference in computer time is dramatic, however. The MCMC solution shown in Fig. 11 was obtained in about 12 hours, while the solution shown in Fig. 15 was completed within some few seconds on a 400 MHz single CPU machine. The difference in the computer time is mainly caused by the

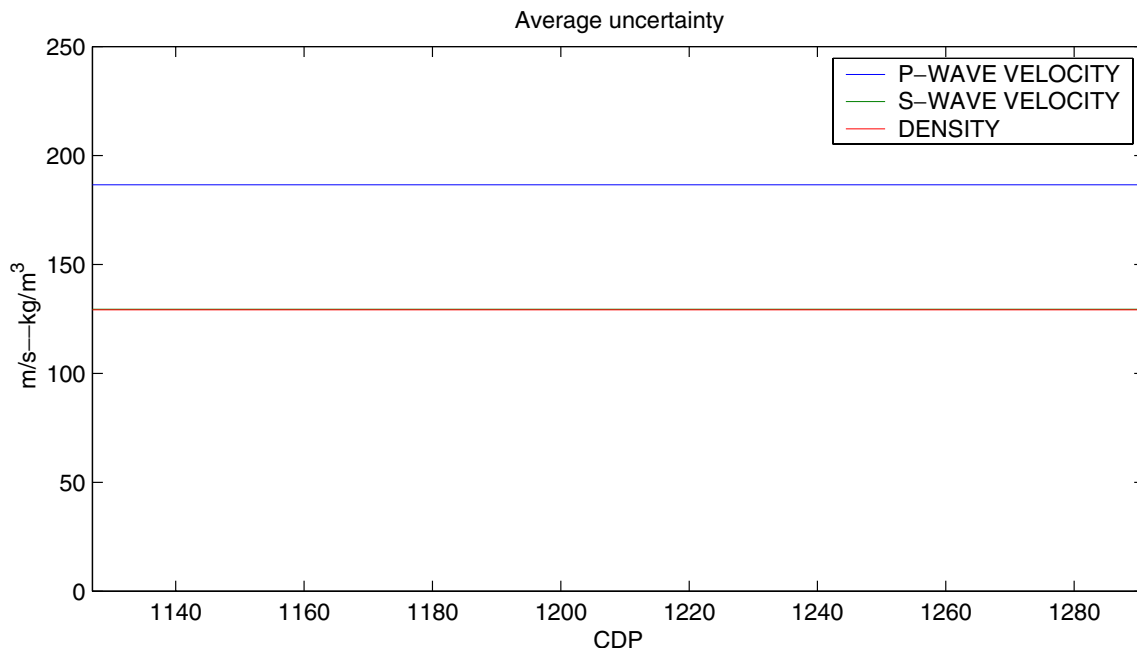


Figure 16 Average posterior uncertainty for P-wave velocity (blue), S-wave velocity (green) and density (red), using fixed wavelet and fixed coloured noise covariance.

500 iterations of the Gibbs' sampler algorithm, but also by the lateral coupling of the model parameters.

CONCLUSIONS

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

ACKNOWLEDGEMENTS

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

REFERENCES

- Aki K. and Richards P.G. 1980. *Quantitative Seismology*. W.H. Freeman & Co.
- Anderson T.W. 1984. *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons Inc.
- Buland A., Kolbjørnsen O. and Omre H. 2003. Rapid spatially coupled AVO inversion in the Fourier domain. *Geophysics* **68**, 824–836.
- Buland A. and Landrø M. 2001. The impact of common offset migration on porosity estimation by AVO inversion. *Geophysics* **66**, 755–762.
- Buland A., Landrø M., Andersen M. and Dahl T. 1996. AVO inversion of Troll Field data. *Geophysics* **61**, 1589–1602.
- Buland A. and Omre H. 2003a. Bayesian linearized AVO inversion. *Geophysics* **68**, 185–198.
- Buland A. and Omre H. 2003b. Bayesian wavelet estimation from seismic and well data. *Geophysics* **68**, in press.
- Carlin B.P. 1996. Hierarchical longitudinal modelling. In: *Markov Chain Monte Carlo in Practice* (eds W.R. Gilks, S. Richardson and D.J. Spiegelhalter), pp. 303–319. Chapman & Hall.
- Chen M., Shao Q. and Ibrahim J. 2000. *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag, Inc.
- Christakos G. 1992. *Random Field Models in Earth Sciences*. Academic Press Inc.

- Duijndam A.J.W. 1988a. Bayesian estimation in seismic inversion. Part I: Principles. *Geophysical Prospecting* **36**, 878–898.
- Duijndam A.J.W. 1988b. Bayesian estimation in seismic inversion. Part II: Uncertainty analysis. *Geophysical Prospecting* **36**, 899–918.
- Geman S. and Geman D. 1984. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.
- Gilks W.R., Richardson S. and Spiegelhalter D.J. 1996. *Markov Chain Monte Carlo in Practice*. Chapman & Hall.
- Gouveia W.P. and Scales J.A. 1998. Bayesian seismic waveform inversion: parameter estimation and uncertainty analysis. *Journal of Geophysical Research* **103**, 2759–2779.
- Hampson D. and Russell B. 1990. AVO inversion: theory and practice. 60th SEG meeting, San Francisco, USA, Expanded Abstracts, 1456–1458.
- Lauritzen S.L. 1996. *Graphical Models*. Oxford University Press.
- Malinverno A. 2000. A Bayesian criterion for simplicity in inverse problem parametrization. *Geophysical Journal International* **140**, 267–285.
- Omre H., Sølna K. and Tjelmeland H. 1993. Simulation of random functions on large lattices. In: *Geostatistics Tróia '92* (ed. A. Soares), pp. 179–199. Kluwer Academic Press.
- Robert C.P. 1994. *The Bayesian Choice*. Springer-Verlag, Inc.
- Scales J.A. and Tenorio L. 2001. Prior information and uncertainty in inverse problems. *Geophysics* **66**, 389–397.
- Smith G.C. and Gidlow P.M. 1987. Weighted stacking for rock property estimation and detection of gas. *Geophysical Prospecting* **35**, 993–1014.
- Spiegelhalter D.J., Best N.G., Gilks W.R. and Inskip H. 1996. Hepatitis B: A case study in MCMC methods. In: *Markov Chain Monte Carlo in Practice* (eds W.R. Gilks, S. Richardson and D.J. Spiegelhalter), pp. 21–43. Chapman & Hall.
- Tarantola A. and Valette B. 1982. Inverse problems = quest for information. *Journal of Geophysics* **50**, 159–170.
- Ulrych T.J., Sacchi M.D. and Woodbury A. 2001. A Bayes tour of inversion: a tutorial. *Geophysics* **66**, 55–69.
- Wang Y. 1999. Simultaneous inversion for model geometry and elastic parameters. *Geophysics* **64**, 182–190.
- Wang Y., White R.E. and Pratt G. 2000. Seismic amplitude inversion for interface geometry: practical approach for application. *Geophysical Journal International* **142**, 162–172.

APPENDIX A

Gaussian distribution

The multi-Gaussian probability density is (Anderson 1984)

$$p(\mathbf{r}) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{r} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{r} - \boldsymbol{\mu})\right], \quad (\text{A1})$$

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

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

$$\begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix} \sim \mathcal{N}_{n_1+n_2} \left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right), \quad (\text{A2})$$

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

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1}(\mathbf{r}_2 - \boldsymbol{\mu}_2), \quad (\text{A3})$$

and covariance,

$$\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}. \quad (\text{A4})$$

APPENDIX B

Gamma and inverse gamma distribution

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

$$p(x|\gamma, \lambda) = \frac{\lambda^\gamma}{\Gamma(\gamma)} x^{\gamma-1} \exp[-\lambda x], \quad (\text{B1})$$

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

$$E\{x|\gamma, \lambda\} = \frac{\gamma}{\lambda}, \quad (\text{B2})$$

$$\text{Var}\{x|\gamma, \lambda\} = \frac{\gamma}{\lambda^2}. \quad (\text{B3})$$

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

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

$$p(x|\gamma, \lambda) = \frac{\lambda^\gamma}{\Gamma(\gamma)} \left(\frac{1}{x}\right)^{\gamma+1} \exp\left[-\frac{\lambda}{x}\right], \quad (\text{B4})$$

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

$$E\{x|\gamma, \lambda\} = \frac{\lambda}{\gamma - 1}, \quad \gamma > 1, \quad (\text{B5})$$

$$\text{Var}\{x|\gamma, \lambda\} = \frac{\lambda^2}{(\gamma - 1)^2(\gamma - 2)}, \quad \gamma > 2. \quad (\text{B6})$$

The expectation and variance in expressions (B5) and (B6) approach infinity when $\gamma \rightarrow 1$ and $\gamma \rightarrow 2$ (from above), respectively. For $\gamma \leq 1$ and $\gamma \leq 2$ ($\gamma > 0$), the expectation and variance are not defined.

For a Gaussian distribution $\mathcal{N}_1(\mu, \sigma^2)$ with unknown variance, the inverse gamma distribution is a conjugate distribution for the variance. That means that if the prior distribution

for the variance is inverse gamma $\sigma^2 \sim \mathcal{IG}(\gamma, \lambda)$, then the posterior distribution is also inverse gamma, but with modified parameters. For x_1, \dots, x_n iid $\mathcal{N}_1(\mu, \sigma^2)$, the posterior distribution for the unknown variance is

$$\sigma^2 | s^2 \sim \mathcal{IG}\left(\gamma + \frac{n}{2}, \lambda + \frac{ns^2}{2}\right), \tag{B7}$$

where

$$s^2 = \sum_{i=1}^n \frac{(x_i - \mu)^2}{n}. \tag{B8}$$

APPENDIX C

Wishart and inverted Wishart distribution

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

$$p(\mathbf{W} | \Sigma, m) = \frac{|\mathbf{W}|^{(m-n-1)/2} \exp\left[-\frac{1}{2}\text{tr}(\Sigma^{-1}\mathbf{W})\right]}{2^{nm/2} |\Sigma|^{m/2} \Gamma_n(m/2)}, \tag{C1}$$

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

$$\Gamma_n(\gamma) = \pi^{n(n-1)/4} \prod_{i=1}^n \Gamma[\gamma - (i-1)/2]. \tag{C2}$$

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

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

$$\mathbf{W} = \sum_{i=1}^m \mathbf{r}_i \mathbf{r}_i^T \sim \mathcal{W}_n(\Sigma, m) \tag{C3}$$

and

$$\mathbf{W}^{-1} \sim \mathcal{IW}_n(\Sigma^{-1}, m), \tag{C4}$$

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

$$E\{\mathbf{W}\} = m\Sigma \tag{C5}$$

and

$$E\{\mathbf{W}^{-1}\} = \frac{1}{m-n-1} \Sigma^{-1}. \tag{C6}$$

Consider Σ to be unknown and stochastic. If Σ is assigned an inverted Wishart prior distribution,

$$\Sigma \sim \mathcal{IW}_n(\Sigma_0, m_0), \tag{C7}$$

then the conditional distribution of Σ , given \mathbf{W} , is

$$\Sigma | \mathbf{W} \sim \mathcal{IW}_n(\Sigma_0 + \mathbf{W}, m_0 + m). \tag{C8}$$

APPENDIX D

The full conditional distributions

The full conditional distribution for an unknown variable conditioned on all the other variables can be constructed only by the terms in expression (16) containing the actual variable. In the following, all the necessary conditional distributions are defined.

The full conditional distribution for \mathbf{m}

From expression (16), the full conditional distribution for the model parameter vector \mathbf{m} can be written

$$p(\mathbf{m} | \cdot) \propto p(\mathbf{d}_{\text{obs}} | \mathbf{m}, \mathbf{s}, \Sigma_{\text{d}}) p(\mathbf{w}_{\text{obs}} | \mathbf{m}, \Sigma_{\text{w}}) p(\mathbf{m} | \boldsymbol{\mu}_{\text{m}}, \Sigma_{\text{m}}), \tag{D1}$$

where the first two terms are the Gaussian likelihoods defined in expressions (4) and (6), and the last term is the Gaussian prior defined in expression (7). Below, the full conditional distribution for \mathbf{m} is calculated by firstly defining $p(\mathbf{d}_{\text{obs}} | \mathbf{s}, \Sigma_{\text{d}})$ and $p(\mathbf{w}_{\text{obs}} | \Sigma_{\text{w}})$. Following Buland and Omre (2003a), let the derivative of $\mathbf{m}(\mathbf{x}, t)$ be

$$\mathbf{m}'(\mathbf{x}, t) = \left[\frac{\partial}{\partial t} \ln \alpha(\mathbf{x}, t), \frac{\partial}{\partial t} \ln \beta(\mathbf{x}, t), \frac{\partial}{\partial t} \ln \rho(\mathbf{x}, t) \right]^T, \tag{D2}$$

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

$$\mathbf{c} = \mathbf{A}\mathbf{m}', \tag{D3}$$

where \mathbf{A} is a sparse matrix defined by the coefficients a_α, a_β and a_ρ in expression (1).

Since \mathbf{m} is assumed to be Gaussian with expectation vector $\boldsymbol{\mu}_{\text{m}}$ and covariance matrix Σ_{m} (expression (7)), then \mathbf{m}' is Gaussian,

$$\mathbf{m}' | \boldsymbol{\mu}_{\text{m}}, \Sigma_{\text{m}} \sim \mathcal{N}_{n_{\text{m}}}(\boldsymbol{\mu}'_{\text{m}}, \Sigma''_{\text{m}}) \tag{D4}$$

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

$$E\{\mathbf{m}'(\mathbf{x}, t)\} = \frac{\partial}{\partial t} \boldsymbol{\mu}_{\text{m}}(\mathbf{x}, t), \tag{D5}$$

and the covariance matrix Σ'_m is defined by

$$\text{Cov}\{\mathbf{m}'(\mathbf{x}, t), \mathbf{m}'(\mathbf{y}, s)\} = \frac{\partial^2}{\partial t \partial s} \Sigma_m(\mathbf{x}, t; \mathbf{y}, s). \quad (\text{D6})$$

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

$$\text{Cov}\{\mathbf{m}'(\mathbf{x}, t), \mathbf{m}(\mathbf{y}, s)\} = \frac{\partial}{\partial t} \Sigma_m(\mathbf{x}, t; \mathbf{y}, s). \quad (\text{D7})$$

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

$$\mathbf{w}_{\text{obs}} = \mathbf{P}\mathbf{m} + \mathbf{e}_w, \quad (\text{D8})$$

and the convolutional model

$$\mathbf{d}_{\text{obs}} = \mathbf{S}\mathbf{A}\mathbf{m}' + \mathbf{e}_d, \quad (\text{D9})$$

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

$$\begin{bmatrix} \mathbf{m} \\ \mathbf{o} \end{bmatrix} \Big| \mathbf{s}, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m, \boldsymbol{\Sigma}_d, \boldsymbol{\Sigma}_w \sim \mathcal{N}_{n_m+n_w+n_d} \left(\begin{bmatrix} \boldsymbol{\mu}_m \\ \boldsymbol{\mu}_o \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_m & \boldsymbol{\Sigma}_{m,o} \\ \boldsymbol{\Sigma}_{o,m} & \boldsymbol{\Sigma}_o \end{bmatrix} \right), \quad (\text{D10})$$

where

$$\boldsymbol{\mu}_o = \begin{bmatrix} \mathbf{P}\boldsymbol{\mu}_m \\ \mathbf{S}\mathbf{A}\boldsymbol{\mu}'_m \end{bmatrix}, \quad (\text{D11})$$

$$\boldsymbol{\Sigma}_o = \begin{bmatrix} \mathbf{P}\boldsymbol{\Sigma}_m\mathbf{P}^T + \boldsymbol{\Sigma}_w & \mathbf{P}\boldsymbol{\Sigma}_m^T\mathbf{A}^T\mathbf{S}^T \\ \mathbf{S}\mathbf{A}\boldsymbol{\Sigma}'_m\mathbf{P}^T & \mathbf{S}\mathbf{A}\boldsymbol{\Sigma}''_m\mathbf{A}^T\mathbf{S}^T + \boldsymbol{\Sigma}_d \end{bmatrix}, \quad (\text{D12})$$

$$\boldsymbol{\Sigma}_{o,m} = \begin{bmatrix} \mathbf{P}\boldsymbol{\Sigma}_m \\ \mathbf{S}\mathbf{A}\boldsymbol{\Sigma}'_m \end{bmatrix}, \quad (\text{D13})$$

and

$$\boldsymbol{\Sigma}_{m,o} = \boldsymbol{\Sigma}_{o,m}^T. \quad (\text{D14})$$

The full conditional distribution for \mathbf{m} is

$$\mathbf{m} \mid \mathbf{o}, \mathbf{s}, \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m, \boldsymbol{\Sigma}_d, \boldsymbol{\Sigma}_w \sim \mathcal{N}_{n_m}(\boldsymbol{\mu}_{m|\cdot}, \boldsymbol{\Sigma}_{m|\cdot}), \quad (\text{D15})$$

where the conditional expectation and covariance are

$$\boldsymbol{\mu}_{m|\cdot} = \boldsymbol{\mu}_m + \boldsymbol{\Sigma}_{m,o}\boldsymbol{\Sigma}_o^{-1}(\mathbf{o} - \boldsymbol{\mu}_o), \quad (\text{D16})$$

$$\boldsymbol{\Sigma}_{m|\cdot} = \boldsymbol{\Sigma}_m - \boldsymbol{\Sigma}_{m,o}\boldsymbol{\Sigma}_o^{-1}\boldsymbol{\Sigma}_{o,m}. \quad (\text{D17})$$

For real-size problems, the dimension of \mathbf{m} is usually large, and an efficient numerical technique is required. For model parameters and seismic data sampled on a regular grid, the spatially

coupled inversion problem can be solved efficiently and exactly in the Fourier domain when the involved covariance matrices are homogeneous and stationary (Buland, Kolbjørnsen and Omre 2003). For more general cases, an approximative solution can be obtained by the sequential screening algorithm (Omre, Sølva and Tjelmeland 1993) (see Appendix E).

The full conditional distribution for s

From expression (16), the full conditional distribution for the wavelet s can be written

$$p(s \mid \cdot) \propto p(\mathbf{d}_{\text{obs}} \mid \mathbf{m}, \mathbf{s}, \boldsymbol{\Sigma}_d) p(\mathbf{s} \mid \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s), \quad (\text{D18})$$

where the first term is the Gaussian likelihood defined in expression (4) and the last term is the Gaussian prior defined in expression (8). Firstly, $p(\mathbf{d}_{\text{obs}} \mid \mathbf{m}, \boldsymbol{\Sigma}_d)$ is calculated. The convolutional model can be written

$$\mathbf{d}_{\text{obs}} = \mathbf{C}\mathbf{s} + \mathbf{e}_d, \quad (\text{D19})$$

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

$$\begin{bmatrix} \mathbf{s} \\ \mathbf{d}_{\text{obs}} \end{bmatrix} \Big| \mathbf{m}, \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \boldsymbol{\Sigma}_d \sim \mathcal{N}_{n_s+n_d} \left(\begin{bmatrix} \boldsymbol{\mu}_s \\ \mathbf{C}\boldsymbol{\mu}_s \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_s & \boldsymbol{\Sigma}_s\mathbf{C}^T \\ \mathbf{C}\boldsymbol{\Sigma}_s & \mathbf{C}\boldsymbol{\Sigma}_s\mathbf{C}^T + \boldsymbol{\Sigma}_d \end{bmatrix} \right). \quad (\text{D20})$$

The full conditional distribution for \mathbf{s} is then

$$\mathbf{s} \mid \mathbf{d}_{\text{obs}}, \mathbf{m}, \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \boldsymbol{\Sigma}_d \sim \mathcal{N}_{n_s}(\boldsymbol{\mu}_{s|\cdot}, \boldsymbol{\Sigma}_{s|\cdot}), \quad (\text{D21})$$

where the conditional expectation and covariance are

$$\boldsymbol{\mu}_{s|\cdot} = \boldsymbol{\mu}_s + \boldsymbol{\Sigma}_s\mathbf{C}^T(\mathbf{C}\boldsymbol{\Sigma}_s\mathbf{C}^T + \boldsymbol{\Sigma}_d)^{-1}(\mathbf{d}_{\text{obs}} - \mathbf{C}\boldsymbol{\mu}_s), \quad (\text{D22})$$

$$\boldsymbol{\Sigma}_{s|\cdot} = \boldsymbol{\Sigma}_s - \boldsymbol{\Sigma}_s\mathbf{C}^T(\mathbf{C}\boldsymbol{\Sigma}_s\mathbf{C}^T + \boldsymbol{\Sigma}_d)^{-1}\mathbf{C}\boldsymbol{\Sigma}_s. \quad (\text{D23})$$

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

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

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

$$\begin{bmatrix} \boldsymbol{\mu}_s \\ \mathbf{s} \end{bmatrix} \Big| \boldsymbol{\Sigma}_s \sim \mathcal{N}_{n_s+n_s} \left(\begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{\mu}_s} \\ \boldsymbol{\mu}_{\boldsymbol{\mu}_s} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\mu}_s} & \boldsymbol{\Sigma}_{\boldsymbol{\mu}_s} \\ \boldsymbol{\Sigma}_{\boldsymbol{\mu}_s} & \boldsymbol{\Sigma}_{\boldsymbol{\mu}_s} + \boldsymbol{\Sigma}_s \end{bmatrix} \right). \quad (\text{D24})$$

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

$$\boldsymbol{\mu}_s | \mathbf{s}, \boldsymbol{\Sigma}_s \sim \mathcal{N}_{n_s} \left(\boldsymbol{\mu}_{\boldsymbol{\mu}_s | \cdot}, \boldsymbol{\Sigma}_{\boldsymbol{\mu}_s | \cdot} \right), \quad (\text{D25})$$

where

$$\boldsymbol{\mu}_{\boldsymbol{\mu}_s | \cdot} = \boldsymbol{\mu}_{\boldsymbol{\mu}_s} + \boldsymbol{\Sigma}_{\boldsymbol{\mu}_s} \left(\boldsymbol{\Sigma}_{\boldsymbol{\mu}_s} + \boldsymbol{\Sigma}_s \right)^{-1} \left(\mathbf{s} - \boldsymbol{\mu}_{\boldsymbol{\mu}_s} \right), \quad (\text{D26})$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\mu}_s | \cdot} = \boldsymbol{\Sigma}_{\boldsymbol{\mu}_s} - \boldsymbol{\Sigma}_{\boldsymbol{\mu}_s} \left(\boldsymbol{\Sigma}_{\boldsymbol{\mu}_s} + \boldsymbol{\Sigma}_s \right)^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\mu}_s}. \quad (\text{D27})$$

The full conditional distribution for σ^2

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

$$\sigma^2 \sim \mathcal{IG}(\gamma, \lambda), \quad (\text{D28})$$

then the full conditional distribution for σ^2 is

$$\sigma^2 | s^2 \sim \mathcal{IG} \left(\gamma + \frac{n}{2}, \lambda + \frac{ns^2}{2} \right), \quad (\text{D29})$$

where s^2 is the sample variance, for example,

$$s_s^2 = \frac{1}{n_s} (\mathbf{s} - \boldsymbol{\mu}_s)^T \boldsymbol{\Sigma}_{0,s}^{-1} (\mathbf{s} - \boldsymbol{\mu}_s), \quad (\text{D30})$$

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

APPENDIX E

The sequential screening algorithm

The full conditional distribution for the complete spatially coupled model parameter vector \mathbf{m} , given by expressions (D15)–(D17), can be factorized as

$$\begin{aligned} p(\mathbf{m} | \cdot) &= p(\mathbf{m}(\mathbf{x}_1) | \cdot) \\ &\times p(\mathbf{m}(\mathbf{x}_2) | \mathbf{m}(\mathbf{x}_1), \cdot) \\ &\vdots \\ &\times p(\mathbf{m}(\mathbf{x}_n) | \mathbf{m}(\mathbf{x}_1), \dots, \mathbf{m}(\mathbf{x}_{n-1}), \cdot), \end{aligned} \quad (\text{E1})$$

where \mathbf{x}_1 to \mathbf{x}_n are the involved surface locations for the inversion in some arbitrary order, and $\mathbf{m}(\mathbf{x}_i)$ is a subvector of \mathbf{m} containing the model parameters at location \mathbf{x}_i . Expression (E1) defines an exact sequential simulation algorithm, but the dimensions of the pdfs on the right-hand side increase rapidly, and the algorithm cannot be used in real-size problems.

The sequential screening algorithm is an approximate version of the exact sequential simulation algorithm defined above. The sequential screening algorithm is based on the assumption that the solution at a location \mathbf{x}_i depends only on a subset of the model and data. The full conditional distribution for $\mathbf{m}(\mathbf{x}_i)$ is obtained by conditioning to a subset of $\{\mathbf{m}(\mathbf{x}_1), \dots, \mathbf{m}(\mathbf{x}_{i-1})\}$, defined by nearest neighbours to \mathbf{x}_i in specified directions, and a subset of \mathbf{w}_{obs} and \mathbf{d}_{obs} . The locations of the wells should define the first locations. Next, the inversion should be run for the boundary locations of the survey, and then a halving procedure is used to select new inversion locations (see Omre *et al.* 1993).