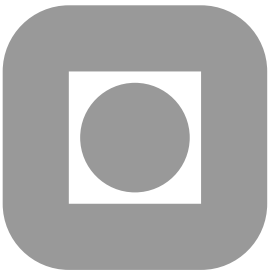
NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

Bayesian wavelet estimation from seisimic and well data

by

Arild Buland and Henning Omre

PREPRINT
STATISTICS NO. 8/2001
ISSN: 0804-9173



NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL http://www.math.ntnu.no/preprint/statistics/2001/S8-2001.ps
Henning Omre has homepage: http://www.math.ntnu.no/∼omre
E-mail: omre@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology,
N-7491 Trondheim, Norway.

Bayesian wavelet estimation from seismic and well data

Arild Buland* and **Henning Omre**†

(March 5, 2001)

ABSTRACT

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

*Statoil Research Centre, Postuttak, N-7005 Trondheim, Norway, and Norwegian University of Science and Technology, N-7491 Trondheim, Norway. E-mail: abu@statoil.com

†Norwegian University of Science and Technology, N-7491 Trondheim, Norway. E-mail: omre@math.ntnu.no

INTRODUCTION

The seismic wavelet is a required input to many seismic processing, modeling and inversion algorithms, and a misspecified wavelet may lead to erroneous results. Uncertainty bounds on the provided wavelet may contribute to realistic uncertainty models related to detailed seismic analysis. By quantifying the uncertainty, for example by including confidence bounds on inversion results, see e.g. Buland and Omre (2001), more robust interpretations of the reservoir characteristics can be made.

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

$$d_{obs}(t) = \int_{-\infty}^{\infty} c(\tau)s(t-\tau)d\tau + e_d(t), \tag{1}$$

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

An obviously complicating factor in the wavelet estimation is a possible time varying mistie between well log reflectivity and the seismic data. In Danielson and Karlsson (1984) it was concluded that such mistie may have serious effects on the wavelet estimation. If uncertainty in the well position and measurement error in the well logs are included, the true reflectivity can be represented by

$$c(t) = c_{obs}(t + \Delta_t(t)) + e_c(t), \tag{2}$$

where $\Delta_t(t)$ and $e_c(t)$ are error terms. The time error function $\Delta_t(t)$ allows for shift, stretch and squeeze of the time axis, while $e_c(t)$ represents errors in the reflection coefficient values. These are related to measurement uncertainty, but also to a possible change in the reflectivity related to uncertainty of the well position relative to the seismic data.

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

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

METHODOLOGY

The methodology will be described in a discrete setting. A discrete version of the convolution model in equation (1) can be written

$$\boldsymbol{d}_{obs} = \boldsymbol{c} \star \boldsymbol{s} + \boldsymbol{e}_d, \tag{3}$$

where $\boldsymbol{d}_{obs}$ represents the seismic data, $\boldsymbol{c}$ is the reflectivity, $\boldsymbol{s}$ is the source wavelet and $\boldsymbol{e}_d$ is the error term. If the vectors $\boldsymbol{d}_{obs}$, $\boldsymbol{c}$, $\boldsymbol{s}$ and $\boldsymbol{e}_d$ represent several traces, for example different offsets or reflection angles, the same notation is used where the convolution is applied trace by trace. The discrete representation of equation (2) is

$$\boldsymbol{c} = \boldsymbol{c}_{obs}(\boldsymbol{\Delta}_t) + \boldsymbol{e}_c, \tag{4}$$

where $\boldsymbol{c}_{obs}(\boldsymbol{\Delta}_t)$ is a discrete representation of the well log reflectivity after modification of the time axis by $\boldsymbol{\Delta}_t$. In practice, the time error may be defined from a set of well markers, $\boldsymbol{t}_{wm} = [t_1, t_2, \ldots, t_m]^T$, which may correspond to major changes in the well logs or interfaces between major layers in a layered subsurface model. A time shift in one of the elements in $\boldsymbol{t}_{wm}$ then corresponds

to a change in the layer thicknesses (in time) of the layers just above and below this marker.

Combining equation (3) and (4) gives the generalized convolution model

$$\boldsymbol{d}_{obs} = \{\boldsymbol{c}_{obs}(\boldsymbol{\Delta}_t) + \boldsymbol{e}_c\} \star \boldsymbol{s} + \boldsymbol{e}_d. \tag{5}$$

The error term $\boldsymbol{e}_c \star \boldsymbol{s}$ is related to well log errors, but components of the seismic error term may have a similar form, $\boldsymbol{e}_d = \boldsymbol{e}_{d_1} \star \boldsymbol{s} + \boldsymbol{e}_{d_2}$, where the first term corresponds to source generated noise, for example remaining multiples. These two error terms are not separable without additional information.

Stochastic model

The problem will now be cast in a Bayesian framework, such that the relevant statistical properties of the problem can be explored. The variables and the problem structure are displayed in Figure 1 as a directed acyclic graph (DAG), see for example Spiegelhalter et al. (1996); Lauritzen (1996). The nodes in the graph represent stochastic variables, and $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote expectations and covariances which will be defined later. The arrows represent the causal dependency structure, where thin lines indicate probability dependencies and thick lines represent deterministic relationships. The DAG can be interpreted as a family tree where the parent nodes point to its children. Before any data are observed, the marginal distributions of the parents are independent, but when their child is known, the properties of the parents become dependent. An example is the reflectivity $\boldsymbol{c}$ and the source $\boldsymbol{s}$ which are apriori independent, but become dependent when the observed seismic data are included in the model.

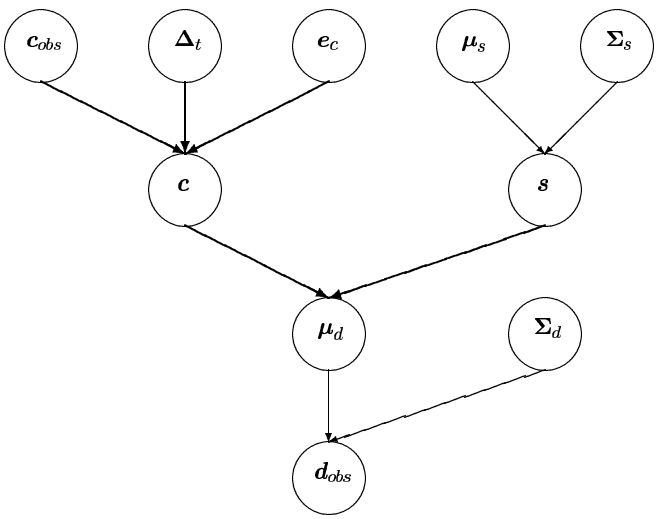


FIG. 1. The convolution model represented by a directed acyclic graph.

When probabilistic dependencies are considered, the deterministic links collapse, so the parents of the seismic data are $\mathrm{Pa}(\boldsymbol{d}_{obs}) = \{\boldsymbol{c}_{obs}, \boldsymbol{\Delta}_t, \boldsymbol{e}_c, \boldsymbol{s}, \boldsymbol{\Sigma}_d\}$, and the parents of the source are $\mathrm{Pa}(\boldsymbol{s}) = \{\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s\}$, where the notation $\mathrm{Pa}(v)$ denotes the parents of a node v. The complete stochastic model can be specified by likelihood terms for the nodes with parents, and prior distributions for the nodes without parents.

Likelihood model

The seismic error term $\boldsymbol{e}_d$ in equation (5) is assumed to be Gaussian with zero expectation and covariance $\boldsymbol{\Sigma}_d$, so the likelihood model for the seismic data is Gaussian,

$$\boldsymbol{d}_{obs} | \boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d \sim \mathcal{N}_{n_d}(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d), \tag{6}$$

where n_d is the dimension, and $\boldsymbol{\mu}_d$ is the expectation defined by the convolution $\boldsymbol{\mu}_d = \boldsymbol{c} \star \boldsymbol{s}$, where $\boldsymbol{c}$ is defined by equation (4). The multi-Gaussian distribution is defined in Appendix A.

The source wavelet is modeled by a Gaussian likelihood,

$$\boldsymbol{s} | \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s \sim \mathcal{N}_{n_s}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s), \tag{7}$$

where n_s is the dimension related to the length of the wavelet, and $\boldsymbol{\mu}_s$ and $\boldsymbol{\Sigma}_s$ are the expectation vector and the covariance matrix, respectively. In Bayesian analysis, the likelihood expression for the source could also be regarded as the first level of an hierarchic prior model, the structural portion, while the distributions of the parameters $\boldsymbol{\mu}_s$ and $\boldsymbol{\Sigma}_s$ form the second level, the subjective portion, see Robert (1994); Carlin (1996).

The length of the wavelet is determined by a compromise between minimizing the data residual and avoiding a too long wavelet. By increasing the wavelet length, the data residual will in general decrease since an increasing amount of local noise is explained by the wavelet. In principle, the wavelet length could be included in the wavelet estimation algorithm as an unknown variable, but this is not done in this study. The source wavelet $\boldsymbol{s}$ may alternatively be represented by a parameterized model, for example Ricker wavelet with unknown amplitude, frequency and time shift, but this is not considered here.

Prior model

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

The prior model for the time error $\boldsymbol{\Delta}_t$ can be defined from a set of well markers. If the well markers can be

restricted to specified intervals, a vague prior model on these intervals is defined by the uniform distribution. If the available prior knowledge is more precise, other distributions may be more appropriate, for example the triangular distribution

$$t_i \sim \mathcal{T}r\,(t_{l_i}, t_{c_i}, t_{u_i}), \tag{8}$$

where the probability is maximum in the center point t_{c_i}, and decreases linearly to zero at t_{l_i} and t_{u_i}. The intervals for the well markers must be defined such that time order for the well markers remain unchanged. This is ensured if the legal time intervals are small compared to the distance between the well markers.

The reflectivity error is assumed to be Gaussian

$$\boldsymbol{e}_c \sim \mathcal{N}_{n_c}(\boldsymbol{0}, \boldsymbol{\Sigma}_c), \tag{9}$$

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

The source wavelet is defined by a Gaussian likelihood with expectation $\boldsymbol{\mu}_s$, and covariance $\boldsymbol{\Sigma}_s$, see expression (7). The uncertainty in these two parameters is defined by prior distributions, ranging from exact values to completely non informative priors. Conjugate prior distributions are often desirable. These are distributions which have the same parametrical prior and posterior forms, but with different parameters. The conjugate distributions for expectation and variance in a Gaussian model are the Gaussian and the inverse gamma distributions, respectively. The expectation vector $\boldsymbol{\mu}_s$ is therefore modeled by a conjugate Gaussian distribution,

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

where $\boldsymbol{\mu}_{\mu_s} = \boldsymbol{0}$, and the variance elements of $\boldsymbol{\Sigma}_{\mu_s}$ are large if the prior knowledge of $\boldsymbol{\mu}_s$ is vague.

If we have sufficient prior knowledge about the structure of the covariance such that the covariance is known up to a multiplicative variance factor, the covariance can be factorized as

$$\boldsymbol{\Sigma}_s = \sigma_s^2 \boldsymbol{\Sigma}_{0_s}, \tag{11}$$

where the uncertainty in the unknown variance factor σ_s^2 is modeled by the inverse gamma distribution

$$\sigma_s^2 \sim \mathcal{IG}(\gamma_s, \lambda_s), \tag{12}$$

which is the conjugate distribution and the standard choice for unknown variances in hierarchic Bayesian models (Robert, 1994). The parameters in expression (12) should be defined in accordance with the prior knowledge. The expectation of the inverse gamma distribution is $\lambda/(\gamma-1)$, valid for $\gamma > 1$, and the variance is $\lambda^2/((\gamma-1)^2(\gamma-2))$, valid for $\gamma > 2$. A precise prior distribution with low variance is obtained by specifying large values for γ and λ such that the expectation becomes correct. A vague prior distribution is obtained by $\gamma \to 1$ and $\lambda \to 0$. See Appendix B for more details about the inverse gamma distribution.

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

$$\boldsymbol{\Sigma}_d = \sigma_d^2 \boldsymbol{\Sigma}_{0_d}, \tag{13}$$

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

$$\sigma_d^2 \sim \mathcal{IG}(\gamma_d, \lambda_d). \tag{14}$$

If the covariance matrices are completely unknown, the factorization in expressions (11) and (13) will not be possible. It is still possible to obtain conjugate prior distributions for the unknown covariances, however, by using the Wishart distribution (Anderson, 1984). This is not used in this study.

Posterior model

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

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

$$p(\boldsymbol{v}|\boldsymbol{d}_{obs}, \boldsymbol{c}_{obs}) = \frac{p(\boldsymbol{v}, \boldsymbol{d}_{obs}, \boldsymbol{c}_{obs})}{p(\boldsymbol{d}_{obs}, \boldsymbol{c}_{obs})} \tag{15}$$

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

$$p(\boldsymbol{v}, \boldsymbol{d}_{obs}, \boldsymbol{c}_{obs}) = \prod_{v_i \in \{\boldsymbol{v}, \boldsymbol{d}_{obs}, \boldsymbol{c}_{obs}\}} p(v_i|\mathrm{Pa}(v_i)). \tag{16}$$

From expressions (15) and (16), and the DAG in Figure 1, the posterior distribution can now be written

$$\begin{aligned} &p(\boldsymbol{\Delta}_t, \boldsymbol{e}_c, \boldsymbol{s}, \boldsymbol{\mu}_s, \sigma_s^2, \sigma_d^2|\boldsymbol{d}_{obs}, \boldsymbol{c}_{obs}) \propto \\ &\quad p(\boldsymbol{d}_{obs}|\boldsymbol{c}_{obs}, \boldsymbol{\Delta}_t, \boldsymbol{e}_c, \boldsymbol{s}, \sigma_d^2) \\ &\quad p(\boldsymbol{\Delta}_t)\, p(\boldsymbol{e}_c)\, p(\boldsymbol{s}|\boldsymbol{\mu}_s, \sigma_s^2)\, p(\boldsymbol{\mu}_s)\, p(\sigma_s^2)\, p(\sigma_d^2), \end{aligned} \tag{17}$$

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

Bayesian inference is completely based on the posterior distribution, where quantities as posterior mean and variance, maximum posterior distribution, quantiles and confidence (credibility) regions may be used. Interesting

quantities can often be expressed as the posterior expectations of a function $f(\boldsymbol{v})$,

$$\mathrm{E}\{f(\boldsymbol{v})|\boldsymbol{d}_{obs},\boldsymbol{c}_{obs}\} = \int f(\boldsymbol{v})\, p(\boldsymbol{v}|\boldsymbol{d}_{obs},\boldsymbol{c}_{obs})\, d\boldsymbol{v}, \quad (18)$$

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

The posterior distribution in expression (17) is not analytically obtainable, but can be explored by stochastic simulation. Several MCMC algorithms are available for this purpose, for example the Gibbs sampler and the Metropolis-Hastings algorithm, see e.g. Gilks et al. (1996); Chen et al. (2000). These algorithms provide a set of samples $\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}, \ldots$ from the posterior distribution $p(\boldsymbol{v}|\boldsymbol{d}_{obs},\boldsymbol{c}_{obs})$. The integral in expression (18) can now be estimated by Monte Carlo integration:

$$\widehat{\mathrm{E}}\{f(\boldsymbol{v})|\boldsymbol{d}_{obs},\boldsymbol{c}_{obs}\} = \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{v}^{(i)}). \quad (19)$$

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

$$[f^{[n\frac{\epsilon}{2}]}, f^{[n(1-\frac{\epsilon}{2})]}], \quad (20)$$

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

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

Gibbs sampler algorithm

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

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

Defining $\boldsymbol{v}_{-j} = \{v_1^{(i)}, \ldots, v_{j-1}^{(i)}, v_{j+1}^{(i-1)}, \ldots, v_n^{(i-1)}\}$, the pseudo-code of the Gibbs sampling algorithm can be written

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

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

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

The conditional distributions are specified in Appendix C.

EXAMPLES

The method is applied on a data set from the Sleipner Øst Field in the Norwegian Block 15/9. Well logs and the corresponding CDP gather are shown in Figure 2, where the CDP gather has been transformed from offset to a set of reflection angles (5°, 13°, 21°, 29°, and 37°). A set of well markers are displayed by horizontal lines, numbered from 1 to 8. The last well marker corresponds to a strong sand/chalk contrast with a reflection coefficient of about 0.3. The average absolute value of the reflection coefficients is 0.02 in the displayed time window.

In the following, wavelet estimation is first done from a CDP stack and zero incidence well log reflectivity, and then from the angle gather in Figure 2 and the corresponding angle dependent well log reflectivity.

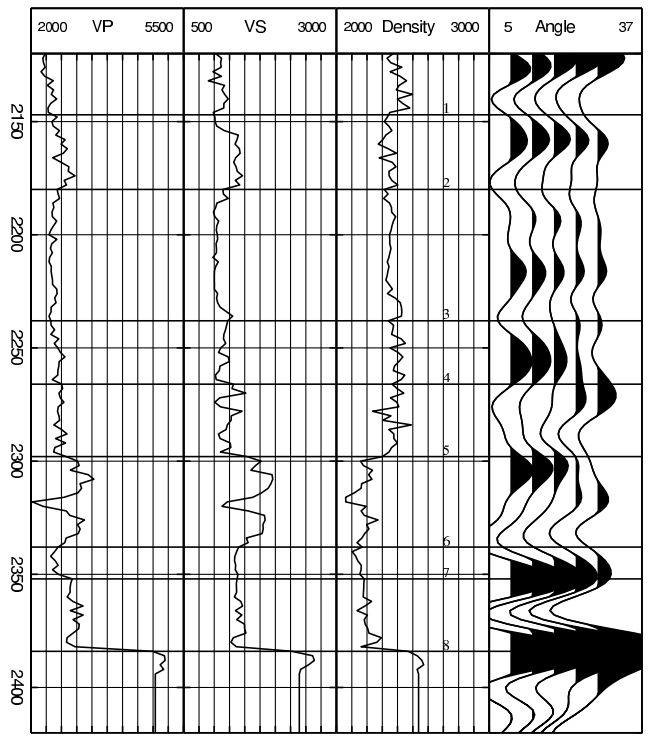


FIG. 2. Well logs and CDP angle gather.

Wavelet estimation from stacked data

As a first test on wavelet estimation from stacked data, assume that there are no errors in the well log information, such that the generalized convolution model in expression (5) reduces to

$$\boldsymbol{d}_{obs} = \boldsymbol{c}_{obs} \star \boldsymbol{s} + \boldsymbol{e}_d. \tag{21}$$

A vague prior model for the unknown source wavelet is defined: The source expectation and covariance have prior distributions

$$\boldsymbol{\mu}_s \sim \mathcal{N}_{n_s}(\mathbf{0}, 10^4 \boldsymbol{\Sigma}_{0_s}), \tag{22}$$
$$\boldsymbol{\Sigma}_s = \sigma_s^2 \boldsymbol{\Sigma}_{0_s}, \tag{23}$$
$$\sigma_s^2 \sim \mathcal{IG}(2,5), \tag{24}$$

where $n_s = 35$, and $\boldsymbol{\Sigma}_{0_s}$ is defined by $\exp[-(|t-t'|/d)^2]$, being a second order exponential correlation function with range d=10 ms, representing the smoothness of the wavelet. The variance of the source expectation in expression (22), $\sigma^2_{\mu_s} = 10^4$, is large compared to the source amplitude. The inverse gamma distributed source variance in expression (24) has expectation equal to 5, and infinite variance.

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

$$\sigma_d^2 \sim \mathcal{IG}(2,5), \tag{25}$$

with expectation equal to 5, and infinite variance.

This first test example has a simple model, but the problem has no trivial analytical solution since both the variance factors σ_s^2 and σ_d^2 are unknown and independent.

The Gibbs sampler algorithm was used to draw samples from the posterior distribution, and 2000 samples were generated. The convergence of the Gibbs simulation can be monitored by plotting variable values for each iteration. A monitor plot of the maximum source amplitude and the seismic noise variance is shown in Figure 3.

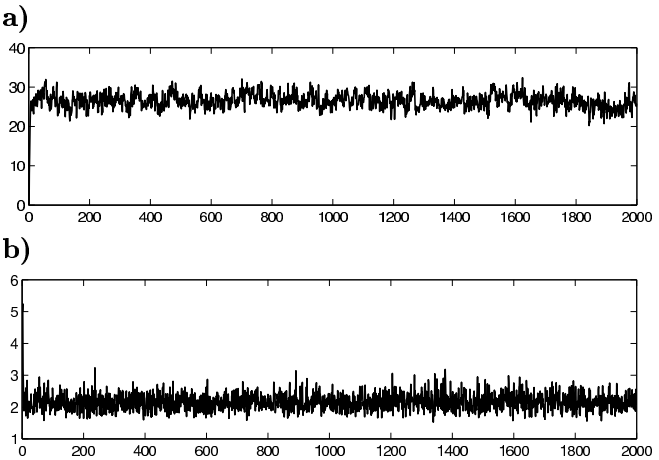


FIG. 3. Monitoring maximum source amplitude (a) and seismic noise variance (b).

The initial values are 0 and 5, respectively, and both quantities converge within some few iterations. Tests with other initial values show the same desirable convergence properties. To be sure, the number of burn-in iterations is set to 100, and these samples are not used in subsequent calculations.

The posterior expectation and 95 % confidence bounds for the wavelet are displayed in Figure 4. The posterior distribution of the unknown seismic noise variance σ_d^2 is shown in Figure 5. The estimated maximum source amplitude is 26.5, and estimated mean data variance is 2.2.

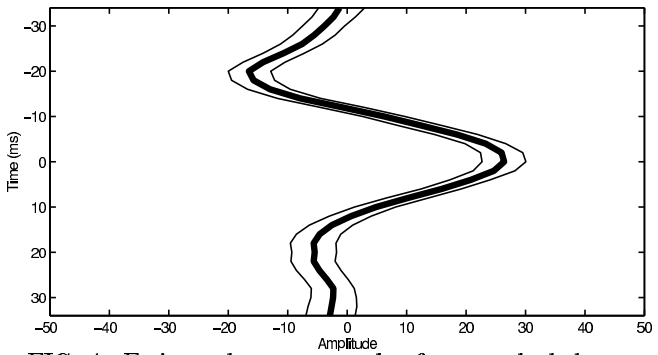


FIG. 4. Estimated source wavelet from stacked data assuming no well log errors.

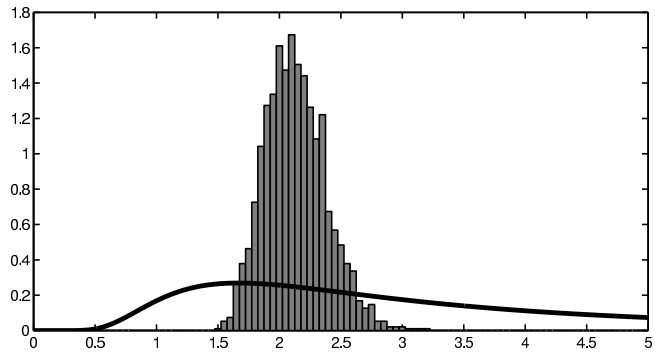


FIG. 5. Estimated seismic noise variance from stacked data assuming no well log errors. The prior model is dashed.

When errors in the well log reflectivity are ignored as in the example above, the estimated seismic noise variance, σ_d^2, is too high. The relatively large misfit between the seismic well log response and the observed seismic data has a damping effect on the wavelet estimation. The error terms $\boldsymbol{\Delta}_t$ and $\boldsymbol{e}_c$ are now included in the model as defined in equation (5). The strong contrast at well marker 8 is used as a reference time, hence well marker 8 is fixed at this reference time to avoid a complete translation of the log. The time shift of the other well markers are triangular distributed

$$t_i \sim \mathcal{T}r\,(-8\,\text{ms}, 0\,\text{ms}, 8\,\text{ms}), \tag{26}$$

and

$$\boldsymbol{e}_c \sim \mathcal{N}_{n_c}(\mathbf{0}, 10^{-5}\boldsymbol{I}). \tag{27}$$

In this case, the posterior expectation and 95% confidence bounds for the wavelet are displayed in Figure 6, while the posterior distribution of the unknown seismic noise variance, σ_d^2, is shown in Figure 7. The estimated maximum source amplitude is now increased to 32.3, while the estimated mean data variance is reduced to 0.26. The posterior distributions of the well markers are shown in Figure 8, where the position of well marker 8 is fixed. The result shows that the total time window of the log is stretched, while the layer between marker 5 and 6 is squeezed.

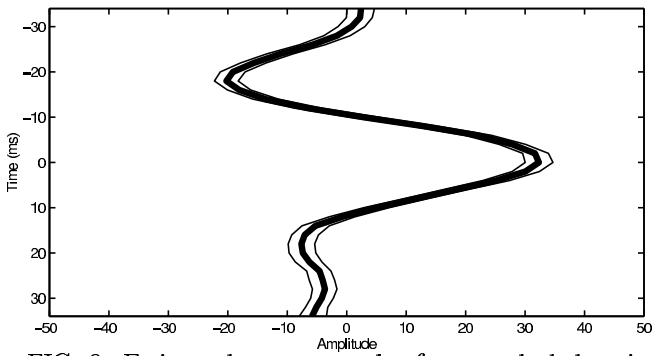


FIG. 6. Estimated source wavelet from stacked data including well log stretching and well log errors.

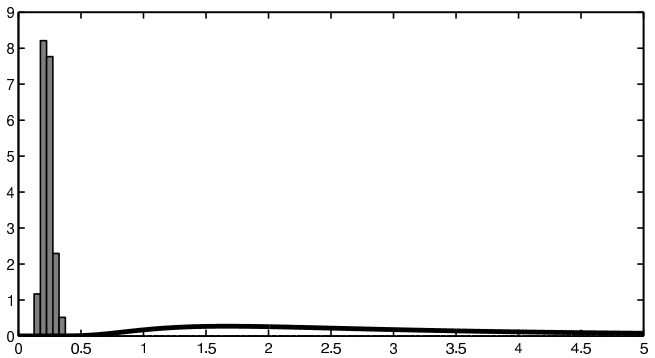


FIG. 7. Estimated seismic noise variance from stacked data including well log stretching and well log errors. The prior model is dashed.

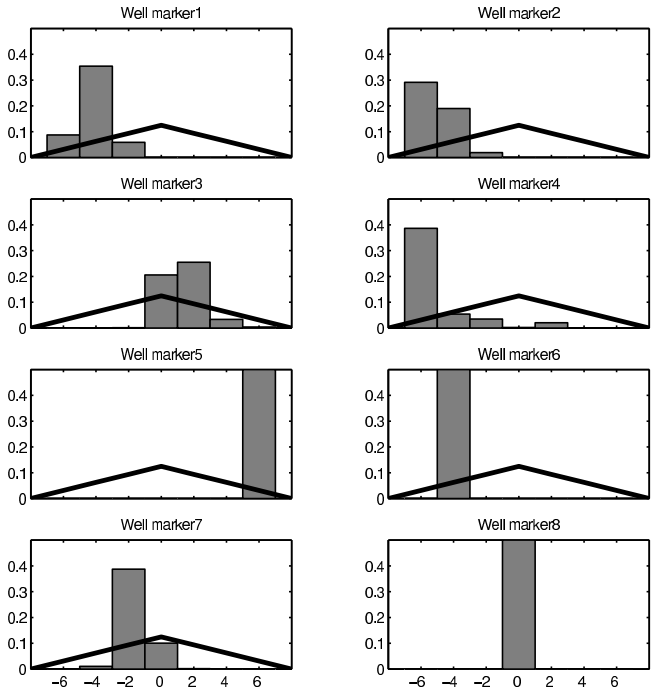


FIG. 8. Estimated time shifts (ms) of the well markers. Note that well marker 8 is fixed. The prior triangular distribution is dashed.

In principle, also the variance of the well log reflectivity error, σ_c^2, could be regarded as an unknown quantity with a vague prior distribution, similarly as σ_s^2 and σ_d^2. The problem with this approach is that the error terms $\boldsymbol{e}_d$ and $\boldsymbol{e}_c$ are not completely separable. The prior model for the variance of the well log error must therefore be specified exactly, or at least relatively precise. In this example, the variance is set to $\sigma_c^2 = 10^{-5}$, such that 95% of the errors are within ± 0.0063, corresponding to two standard deviations from the expectation value. For comparison, the mean well log reflection coefficient is 0.02 (absolute value). The effect of different σ_c^2 is shown in Figure 9. For reflection error variances close up to 10^{-4}, the estimation of the source wavelet is stable, also with respect to uncertainty. Variances larger than 10^{-4} allows for reflectivity errors larger than the typical well log reflection coefficient, and this has a dramatic damping effect on the wavelet. As expected, the estimated seismic noise variance decreases with increasing reflection error variance. Reflection error variances below 10^{-6} have only a marginal effect.

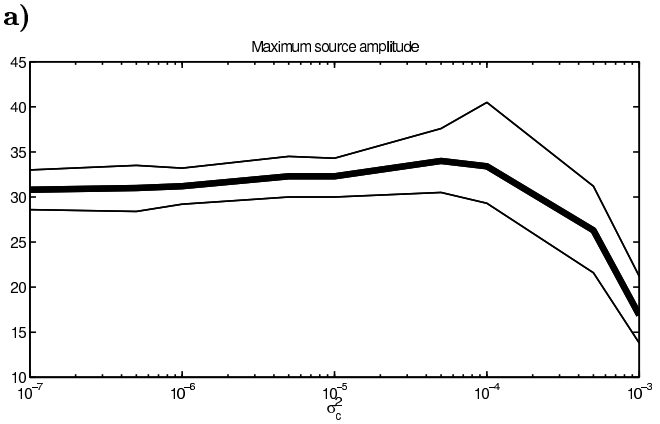


FIG. 9. Maximum source amplitude (a) and seismic noise variance (b) as function of reflectivity error variance, σ_c^2. The thin lines show the 95 % confidence intervals.

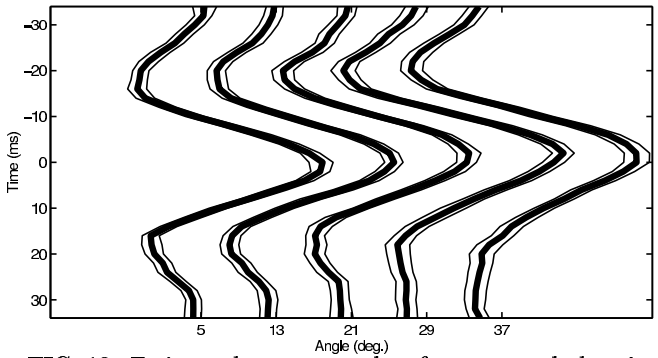


FIG. 10. Estimated source wavelets from prestack data including well log stretching and well log errors.

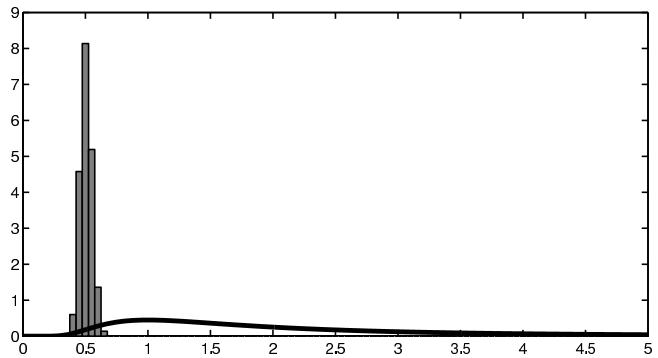


FIG. 11. Estimated seismic noise variance from prestack data including well log stretching and well log errors. The prior model is dashed.

Wavelet estimation from prestack data

The extension of the wavelet estimation to prestack data is straightforward. In this prestack example, a complete model is used including well log errors, and the prior distributions are as in the poststack example. In the simplest case, one wavelet is estimated independently for each reflection angle, but it is reasonable to assume a certain degree of similarity between the wavelets. This is obtained by a correlation function, $\rho_\theta = \exp[-|\theta_i - \theta_j|/d_\theta]$, where $d_\theta = 20°$ is the choosen correlation range.

From Gibbs sampling, the posterior expectation and 95 % confidence bounds for the five angle wavelets are displayed in Figure 10. The maximum amplitudes are slightly higher than the poststack amplitude, ranging from 35.2 to 40.8. The posterior distribution of the unknown seismic noise variance, σ_d^2, is shown in Figure 11. The estimated mean value of the noise variance is 0.52, which is twice the stack noise variance. The posterior distributions of the time shifts are more peaked than in the poststack example, see Figure 12.

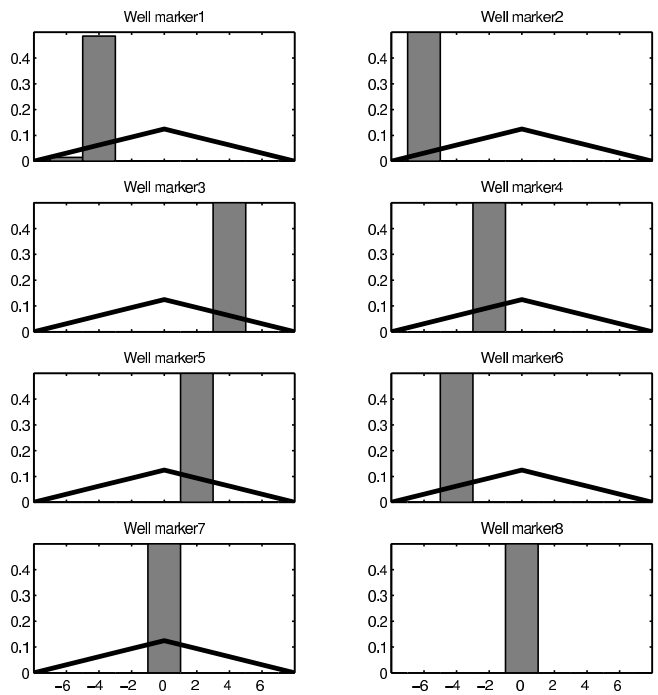


FIG. 12. Estimated time shifts (ms) of the well markers. Note that well marker 8 is fixed. The prior triangular distribution is dashed.

CONCLUSIONS

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

ACKNOWLEDGMENTS

We thank Statoil for permission to publish this paper.

REFERENCES

Anderson, T. W., 1984, An introduction to multivariable statistical analysis: John Wiley and Sons Inc.

Buland, A., and Omre, H., 2001, Bayesian AVO inversion: Submitted to Geophysics, Oct. 2000. Preprint: http://www.math.ntnu.no/preprint/statistics/2001.

Carlin, B. P., 1996, Hierarchical longitudinal modelling, *in* Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., Eds., Markov Chain Monte Carlo in practice: Chapman & Hall, 303–319.

Chen, M., Shao, Q., and Ibrahim, J., 2000, Monte Carlo methods in Bayesian computation: Springer.

Danielson, V., and Karlsson, T. V., 1984, Extraction of signatures from seismic and well data: First Break, **2**, no. 4, 15–21.

Geman, S., and Geman, D., 1984, Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images: IEEE Trans. Pattern. Anal. Mach. Intell., **6**, 721–741.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., 1996, Markov Chain Monte Carlo in practice: Chapman & Hall.

Lauritzen, S. L., 1996, Graphical models: Oxford University Press.

Lines, L. R., and Treitel, S., 1985, Wavelets, well logs and Wiener filters: First Break, **3**, no. 8, 9–14.

Nyman, D. C., Parry, M. J., and Knight, R. D., 1987, Seismic wavelet estimation using well control: 57th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 535–537.

Poggiagliolmi, E., and Allred, R. D., 1994, Detailed reservoir definition by integration of well and 3-D seismic data using space adaptive wavelet processing: The Leading Edge, **13**, no. 7, 749–754.

Richard, V., and Brac, J., 1988, Wavelet analysis using well log information: 58th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 946–949.

Robert, C. P., 1994, The Bayesian choice: Springer-Verlag.

Spiegelhalter, D. J., Best, N. G., Gilks, W. R., and Inskip, H., 1996, Hepatitis B: A case study in MCMC methods, *in* Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., Eds., Markov Chain Monte Carlo in practice: Chapman & Hall, 21–43.

APPENDIX A–GAUSSIAN DISTRIBUTION

The multi-Gaussian probability density is

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

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

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

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

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

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

and covariance

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

APPENDIX B–GAMMA AND INVERSE GAMMA DISTRIBUTION

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

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

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

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

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

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

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

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

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

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

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

For $\gamma \leq 1$ and $\gamma \leq 2$, the expectation and variance in expressions (B-5) and (B-6) are not defined.

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

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

where

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

APPENDIX C–CONDITIONAL DISTRIBUTIONS IN GIBBS SAMPLING

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

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

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

Conditional distribution for $\boldsymbol{\Delta}_t$

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

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

Conditional distribution for $\boldsymbol{e}_c$

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

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

where the matrix $\boldsymbol{S}$ represents the given source wavelet.

The conditional simultaneous distribution for the reflectivity error and the seismic data are

$$\begin{bmatrix} \boldsymbol{e}_c \\ \boldsymbol{d}_{obs} \end{bmatrix} \Big| \boldsymbol{c}_{obs}, \boldsymbol{\Delta}_t, \boldsymbol{s}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_d \Big] \sim \mathcal{N}_{n_c+n_d} \left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{S}\boldsymbol{c}_{obs}(\boldsymbol{\Delta}_t) \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_c & \boldsymbol{\Sigma}_c \boldsymbol{S}^T \\ \boldsymbol{S}\boldsymbol{\Sigma}_c & \boldsymbol{S}\boldsymbol{\Sigma}_c\boldsymbol{S}^T + \boldsymbol{\Sigma}_d \end{bmatrix} \right). \tag{C-3}$$

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

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

where

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

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

Conditional distribution for $\boldsymbol{s}$

The convolution can be reformulated to a matrix-vector multiplication

$$\boldsymbol{d}_{obs} = \boldsymbol{C}\boldsymbol{s} + \boldsymbol{e}_d, \tag{C-7}$$

where the matrix $\boldsymbol{C}$ represents the given reflectivity $\boldsymbol{c} = \boldsymbol{c}_{obs}(\boldsymbol{\Delta}_t) + \boldsymbol{e}_c$.

The conditional simultaneous distribution for the wavelet and the seismic data are

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

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

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

where

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

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

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

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

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

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

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

where

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

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

Conditional distribution for σ_s^2

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

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

then the full conditional distribution for σ_s^2 is

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

where

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

Conditional distribution for σ_d^2

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

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

then the full conditional distribution for σ_d^2 is

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

where

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