On inferring the noise in probabilistic seismic AVO inversion using hierarchical Bayes
Madsen, Rasmus Bødker; Zunino, Andrea; Hansen, Thomas Mejer

Published in:
SEG International Exposition and Annual Meeting, 24-29 September, Houston, Texas

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
On inferring the noise in probabilistic seismic AVO inversion using hierarchical Bayes

Rasmus Bødker Madsen*, Andrea Zanino and Thomas Mejer Hansen
Niels Bohr Institute, University of Copenhagen

SUMMARY

A realistic noise model is essential for trustworthy inversion of geophysical data. Sometimes, as in case of seismic data, quantification of the noise model is non-trivial. To remedy this, a hierarchical Bayes approach can be adopted in which properties of the noise model, such as the amplitude of an assumed uncorrelated Gaussian noise model, can be inferred as part of the inversion. Here we demonstrate how such an approach can lead to substantial overfitting of noise when inverting a 1D reflection seismic NMO data set. We then argue that usually the noise model is correlated, and suggest to infer the amplitude of a correlated Gaussian noise model. This provides better results than assuming an uncorrelated model. In general though, the results suggest that care should be taken using the hierarchical Bayes approach to infer the noise model.

INTRODUCTION

Data \( (d \in \mathbb{R}^N) \) are the unique forward response from a physical model \((g \in \mathbb{R}^M)\):

\[
d = g(m) \tag{1}
\]

In nature, observed data \((d_{\text{obs}})\) are not noise-free, i.e. \(d_{\text{obs}} = g(m) + \varepsilon\). The inverse problem of inferring values of the model parameters from the observed data are therefore non-unique (Sen and Stoffa, 1996; Tarantola, 2005).

The chosen likelihood function, which measures how well the model parameters match the observed data, is dependent on the distribution of noise (Box and Tiao, 1992). The likelihood, and hence the final posterior distribution of model parameters, can be very sensitive to noise-level and noise models (Thore, 2015). The noise model should ideally include information about measurement and experimental errors as well as account for imperfections in the forward model and/or simplifications due to the parametrization (Dosso and Holland, 2006). However, it is often argued that since no theory is exact, all features that are not captured by the theory are just observational errors, and no distinction should be made between different types of noise (Sen and Stoffa, 1996).

In the classical Bayesian stochastic inference paradigm information on the uncertainty distribution of the data should be established a priori as part of the likelihood function. If for instance a Gaussian noise model is assumed, the corresponding Gaussian likelihood takes the following form (Box and Tiao, 1992)

\[
p(d_{\text{obs}}|m) = \exp \left( -\frac{1}{2}(d_{\text{obs}} - g(m))^T C_D (d_{\text{obs}} - g(m)) \right) \tag{2}
\]

where \(C_D\) is the data covariance model. The data covariance can be split into contributions from measurement error \(C_d\) and theory error \(C_T\), assuming independence of the two. (Mosegaard and Tarantola, 2002)

\[
C_D = C_d + C_T. \tag{3}
\]

An extension to the classical paradigm is offered through the use of the hierarchical Bayesian approach, where parameters of the noise (and of the prior) can be inferred from the observed data (Gelman et al., 2014). A hierarchical scheme has been used in various geophysical inverse problems in order to infer information about noise level from data (Buland and Omre, 2003b; Malinverno and Briggs, 2004; Malinverno and Parker, 2006; Bodin et al., 2012; Dettmer and Dosso, 2012; Ray et al., 2013).

Buland and Omre (2003b) explicitly used a hierarchical Bayes formulation of the inverse problem, enabling noise-level estimation as well as wavelet estimation in a joint AVO (Amplitude Versus Offset) inversion scheme. They concluded that estimating the seismic noise model as part of the probabilistic inversion is viable. However, in their real data case only negligible improvements were gained on posterior variance of model parameters, compared to estimating the noise covariance and wavelet prior to AVO inversion. As opposed to Buland and Omre (2003b), we will analyze the posterior resolution (not the posterior variance) obtained using hierarchical Bayes with different assumptions about the noise model. Specifically we will investigate whether the noise is underestimated, which will result in an apparent smaller posterior variability. However, such reduced posterior variability might reflect non-existent features appearing as a result of fitting noise as data. We propose a set of synthetic tests similar to that of Buland and Omre (2003a). The synthetic tests are based on a reference model from which possible posterior biases in the model parameters can be assessed alongside the posterior variance.

THEORY AND METHOD

Forward model and prior information

The synthetic data (AVO gather) are created following the methodology of Buland and Omre (2003a). The problem here is to infer information about the three elastic parameters: P-wave \((v_p)\), S-wave \((v_s)\), and density \((\rho)\) from a seismic AVO gather. The observed data can be expressed as a linear convolution forward problem with some added noise:

\[
d_{\text{obs}} = Gm + \varepsilon = WADm + e \tag{4}
\]

where \(W\) is the wavelet matrix which is convolved with the reflectivity series \(ADm\). The prior distribution of model parameters \(p(m)\) is Gaussian and the three elastic parameters have an internal correlation of 0.7. This prior corresponds to the "Well B" scenario of (Buland and Omre, 2003a). In order to allow...
On inferring noise using hierarchical Bayes

A trickier aspect of noise in seismic AVO data is the fact that the error is often systematic (Riedel et al., 2003). In order to test the hierarchical Bayesian approach of inferring noise, the colored noise is set to:

\[ m = \text{[ln}(v_p/v'_p)^T, \text{ln}(v_s/v'_s)^T, \text{ln}(\rho/\rho'0)^T]^T \sim \text{Gauss}(\mu, \Sigma_M) \]  

(5)

where \( \mu \) is the prior expectation of the model parameters, \( v'_p, v'_s, \rho'0 \) are reference values, and \( \Sigma_M \) describes the prior covariance. The noise on the observed data is \( e = e_{\text{uncor}} + e_{\text{cor}} \), where the noise model is Gaussian and split into two components as in equation 3. Here we let \( \Sigma_d \) be the uncorrelated white noise component and \( \Sigma_T \) be the correlated colored noise component, so that:

\[ e_{\text{uncor}} \sim \text{Gauss}(0, \Sigma_d), \quad e_{\text{cor}} \sim \text{Gauss}(0, \Sigma_T). \]  

(6)

A SNR = 30 is chosen for the uncorrelated noise. This number is perhaps slightly generous towards the filtering processes. On the other hand, white uncorrelated noise is different in wavelength from the seismic signal and can often easily be filtered out during processing of the raw data (Vecken and Da Silva, 2004). Most of the remaining noise would therefore tend to resemble the observed data in the frequency domain.

Bayesian linearized AVO inversion

Since the inverse problem is linear Gaussian, the posterior distribution of the model parameters is assumed to be log-Gaussian, so:

\[ \mathbf{m} = [\text{ln}(v_p/v'_p)^T, \text{ln}(v_s/v'_s)^T, \text{ln}(\rho/\rho'0)^T]^T \sim \text{Gauss}(\mu, \Sigma_M). \]  

(5)

where \( \mu \) is the prior expectation of the model parameters, \( v'_p, v'_s, \rho'0 \) are reference values, and \( \Sigma_M \) describes the prior covariance. The noise on the observed data is \( e = e_{\text{uncor}} + e_{\text{cor}} \), where the noise model is Gaussian and split into two components as in equation 3. Here we let \( \Sigma_d \) be the uncorrelated white noise component and \( \Sigma_T \) be the correlated colored noise component, so that:

\[ e_{\text{uncor}} \sim \text{Gauss}(0, \Sigma_d), \quad e_{\text{cor}} \sim \text{Gauss}(0, \Sigma_T). \]  

(6)

A SNR = 30 is chosen for the uncorrelated noise. This number is perhaps slightly generous towards the filtering processes. On the other hand, white uncorrelated noise is different in wavelength from the seismic signal and can often easily be filtered out during processing of the raw data (Vecken and Da Silva, 2004). Most of the remaining noise would therefore tend to resemble the observed data in the frequency domain.

Bayesian linearized AVO inversion

Since the inverse problem is linear Gaussian, the posterior distribution of the model parameters is assumed to be log-Gaussian, so:

\[ \mathbf{m} = [\text{ln}(v_p/v'_p)^T, \text{ln}(v_s/v'_s)^T, \text{ln}(\rho/\rho'0)^T]^T \sim \text{Gauss}(\mu, \Sigma_M) \]  

(5)

where \( \mu \) is the prior expectation of the model parameters, \( v'_p, v'_s, \rho'0 \) are reference values, and \( \Sigma_M \) describes the prior covariance. The noise on the observed data is \( e = e_{\text{uncor}} + e_{\text{cor}} \), where the noise model is Gaussian and split into two components as in equation 3. Here we let \( \Sigma_d \) be the uncorrelated white noise component and \( \Sigma_T \) be the correlated colored noise component, so that:

\[ e_{\text{uncor}} \sim \text{Gauss}(0, \Sigma_d), \quad e_{\text{cor}} \sim \text{Gauss}(0, \Sigma_T). \]  

(6)

A SNR = 30 is chosen for the uncorrelated noise. This number is perhaps slightly generous towards the filtering processes. On the other hand, white uncorrelated noise is different in wavelength from the seismic signal and can often easily be filtered out during processing of the raw data (Vecken and Da Silva, 2004). Most of the remaining noise would therefore tend to resemble the observed data in the frequency domain.

Bayesian linearized AVO inversion

Since the inverse problem is linear Gaussian, the posterior distribution of the model parameters is assumed to be log-Gaussian, so:

\[ \mathbf{m} = [\text{ln}(v_p/v'_p)^T, \text{ln}(v_s/v'_s)^T, \text{ln}(\rho/\rho'0)^T]^T \sim \text{Gauss}(\mu, \Sigma_M) \]  

(5)

where \( \mu \) is the prior expectation of the model parameters, \( v'_p, v'_s, \rho'0 \) are reference values, and \( \Sigma_M \) describes the prior covariance. The noise on the observed data is \( e = e_{\text{uncor}} + e_{\text{cor}} \), where the noise model is Gaussian and split into two components as in equation 3. Here we let \( \Sigma_d \) be the uncorrelated white noise component and \( \Sigma_T \) be the correlated colored noise component, so that:

\[ e_{\text{uncor}} \sim \text{Gauss}(0, \Sigma_d), \quad e_{\text{cor}} \sim \text{Gauss}(0, \Sigma_T). \]  

(6)
On inferring noise using hierarchical Bayes

Figure 1: Posterior distribution of $p(\mathbf{m|m_{obs}})$ and $p(\mathbf{h|m_{obs}})$, log-likelihood and cross-correlation for case 1 of hierarchical Bayes inversion using $C_D = h_1^2 C_d$

Figure 2: Histograms of the hyperposterior distributions for all MCMC runs $p(\mathbf{h|m_{obs}})$. Notice the logarithmic scale used for Case 1 (top plot).

overfitting of the posterior for the model parameters and the algorithms ability to correctly determine the variance of uncorrelated noise on the data while underestimating the actual total noise level.

The correlation between adjacent samples is not high, as the correlation decreases to the average level quickly form the last sample. The pattern of log-likelihood and cross-correlation is similar for all the following MCMC runs in general and are therefore not shown.

The overfitting of the hierarchical Bayes inversion with the uncorrelated noise model becomes even more apparent in Figure 3. Here the percentage of the reference model being inside the confidence interval is plotted as a function of the size of the confidence interval. It is clearly visible that the posterior distribution of the elastic variables is not capturing the reference model for all confidence intervals. The apparently small posterior variability is actually reflecting non-existent features, i.e. noise being fitted as data. The result of the hierarchical inversion is similar to just applying a linear AVO inversion with $\sigma_d = 0.0039$. To evaluate the consistency of the results from case 1, the algorithm is run an additional time. The histogram of the hyperposterior of $h_1$ in Figure 2 for the secondary run shows the same pattern as for the first run. This indicates a certain level of consistency in the MCMC results.

Hierarchical Bayes inversion - Case 2

In the second test case we assume a known shape of colored noise $C_{T,\text{shape}}$, i.e. we test whether it is possible at all to infer the variance of the noise knowing the reference color. The standard deviation of the colored noise is set as an additional hyperparameter $\sigma_T = h_2$. The noise model then takes the following form: $C_D = h_1^2 C_{d,\text{shape}} + h_2^2 C_{T,\text{shape}}$. Again, we set a rather broad uniform hyperprior distribution for the standard deviation of the correlated noise: $p(h_2) \sim \text{Unif}(0.0001, 1)$. The result for two runs are summarized in Figure 2 and 3. Both MCMC runs show approximately the same hyperposterior dis-
On inferring noise using hierarc. Bayes

Figure 3: Percentage of reference model being inside the confidence interval, i.e. non-yellow areas in Figure 1), as a function of confidence interval. The uppermost figures are calculated using Bayesian linearized AVO inversion for reference.

Hierarchical Bayes inversion - Case 3

Since the correct shape of the noise is never readily available in a real-world scenario, we propose to estimate or assume some correlation of the noise prior to inversion. We assume that the noise is showing smoothness comparable with the wavelet. Furthermore, the noise is correlated between the individual angle-stacks. The correlation between angle stacks is believed to vary slightly as a function of angle. This is set up in the angle-stacks. The correlation between angle stacks is believed the noise is showing smoothness comparable with the wavelet.

Since the correct shape of the noise is never readily available in a real-world scenario, we propose to estimate or assume some correlation of the noise prior to inversion. We assume that the noise is showing smoothness comparable with the wavelet. Furthermore, the noise is correlated between the individual angle-stacks. The correlation between angle stacks is believed to vary slightly as a function of angle. This is set up in the angle-stacks. The correlation between angle stacks is believed the noise is showing smoothness comparable with the wavelet.

The variance of the correlated noise $b_2$ (red dot) is slightly overestimated, whereas the variance of the uncorrelated noise $b_1$ is well-determined. The nearly-correct estimation of variance level, results in the reference model being within the confidence interval following the ideal line in Figure 3. The model therefore appears to neither over- nor underfit the data. The results are similar to doing a linear AVO inversion with the reference noise model.

Hierarchical Bayes inversion - Case 3

Since the correct shape of the noise is never readily available in a real-world scenario, we propose to estimate or assume some correlation of the noise prior to inversion. We assume that the noise is showing smoothness comparable with the wavelet. Furthermore, the noise is correlated between the individual angle-stacks. The correlation between angle stacks is believed to vary slightly as a function of angle. This is set up in the following manner

$$W_{\text{corr}} = \begin{bmatrix} \beta_1 I & \beta_2 \beta_2 I & \ldots & \beta_n \beta_n I \\ \beta_1 \beta_1 I & \beta_2 I & \ldots & \beta_n \beta_n I \\ \vdots & \vdots & \ddots & \vdots \\ \beta_1 \beta_1 I & \beta_2 \beta_2 I & \ldots & \beta_n I \end{bmatrix}$$

where $I$ is the identity matrix with the size of one individual angle stack. A simple model can easily be derived by setting $\beta = [\beta_1, \beta_2, \ldots, \beta_n] = 1$. Using the variance of each angle stack on the observed data offers an estimate for $\beta$. These beta values could possibly also be obtained using optimization of the marginal likelihood in equation 11. In our case we use $\beta = [1, 0.99, 0.98, 0.95, 0.92, 0.90, 0.88]$. The final approximate shape of the colored noise is then:

$$C_{W,\text{shape}} = \frac{W W_{\text{corr}} W^T}{\max |W W_{\text{corr}} W|}$$

Using the same approach as for case 2, the approximate noise model takes the following form: $C_D = \beta_1^2 C_{d,\text{shape}} + \beta_2^2 C_{W,\text{shape}}$. The variance of the uncorrelated noise is as for case 1 and 2 correctly estimated as both histograms are centered around the correct value (black dot) in Figure 2. However, the variance of the uncorrelated noise is underestimated when comparing the hyperprior distribution of $p(h_1|d_{\text{obs}})$ with the correct value (blue dot). As for both case 1 and case 2, there is consistency between the results from the two MCMC runs. Figure 3 shows improvements of case 3 compared to case 1 for all elastic parameters but the $V_p/V_s$ ratio, which is still not captured by the posterior distribution. This indicates that overfitting of the data is still present using the approximate shape, but is nevertheless reduced compared to simply assuming an uncorrelated shape of the noise.

CONCLUSION

Our results in general indicate that caution should be taken when inferring the noise as an additional parameter in inversion. It seems that the assumption of uncorrelated noise in case 1 is not good for inferring the correct noise level on data with correlated noise. The hierarchical Bayes approach was in all cases able to accurately estimate the variance of the uncorrelated noise on the data. However, using the approach of case 1 the total variance of the noise is not recovered and significant overfitting of the data was demonstrated. Choosing the correct shape of the noise, as in case 2, eliminates the overfitting, and a correct variance is estimated even for a wide hyperprior. The results from case 2 indicate that it is possible to estimate a correct variance of the noise model using the data. Finally, for case 3 with an approximate shape of the noise, the variance estimate of the noise is improved compared to case 1. The model is however still overfitting the data. Especially $V_p/V_s$ shows an apparent smaller posterior variability that is not capturing the true model. In a real world case it is probably reasonable to assume that substantial knowledge about the general noise-level is available to constrain the wide hyperprior distribution (Gelman et al., 2014). This could potentially improve the result for an approximate noise model. For our synthetic tests, using an approximate shape of the correlated noise offers an improvement on the noise-level estimate compared with using an uncorrelated noise model. Further work could nonetheless be done to obtain more reliable estimates for an approximate shape of the noise, which could further improve the posterior resolution of the hierarchical Bayes approach in general.
REFERENCES
Note: This reference list is a copyedited version of the reference list submitted by the author. Reference lists for the 2017 SEG Technical Program Expanded Abstracts have been copyedited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES


