

# Nonlinear Bayesian joint inversion of seismic reflection coefficients

Tor Erik Rabben<sup>1,\*</sup>, Håkon Tjelmeland<sup>2</sup>, and Bjørn Ursin<sup>1</sup>

<sup>1</sup> *Department of Petroleum Engineering and Applied Geophysics, Norwegian University of Science and Technology, 7491 Trondheim, Norway*

<sup>2</sup> *Department of Mathematical Sciences, Norwegian University of Science and Technology, 7491 Trondheim, Norway*

\* *E-mail: torerik@ipt.ntnu.no*

Accepted 1999 November 11. Received 1999 October 6; in original form 2007 February 9

## SUMMARY

Inversion of the seismic reflection coefficients are formulated in a Bayesian framework. Measured reflection coefficients and model parameters are assigned statistical distributions based on information known prior to the inversion, and together with the forward model uncertainties are propagated into the final result. This enables a quantification of the reliability of the inversion. A quadratic approximation to the Zoeppritz equations is used as the forward model. Compared with the linear approximation the bias is reduced and the uncertainty estimates are more reliable. The differences when using the quadratic approximations and the exact expressions are minor. The solution algorithm is sampling based and because of the nonlinear forward model the Metropolis-Hastings algorithm is used. To achieve convergence it is important to keep strict control of the accept probability in the algorithm. Joint inversion using information from both reflected PP-waves and converted PS-waves yield smaller bias and reduced uncertainties compared to using only reflected PP-waves.

**Key words:** Seismic inversion – statistical methods

## 1 INTRODUCTION

The seismic reflection coefficients contain information about elastic parameters in the subsurface. In an amplitude versus angle (AVA) inversion, the main objective is to estimate elastic parameters from the reflection coefficients. Elastic parameters can in turn be used for sediment classification and extraction of fluid properties. Given a model it is easy to find an answer to the inverse problem, the real challenge is to find a *reasonable* answer (Hampson, 1991).

Mathematically, the main problem is nonuniqueness. Analytical expressions for the reflection coefficients are given by the Zoeppritz equations. These equations are complicated and highly nonlinear. In the situation of reflections between two isotropic media the equations involves 5 parameters. Ursin & Tjøland (1996) showed that in practice only up to three parameters can be estimated from pre-critical PP reflection coefficients. To help overcome this, linear approximations to the Zoeppritz equations have been derived (e.g. Aki & Richards (1980) and Smith & Gidlow (1987)), both for their simplicity and the favourable reduction from 5 to 3 parameters. One other possible solution is to include PS reflections in the inversion. Because pressure waves and shear waves sense different rock and pore-fluid properties, joint PP and PS data can provide superior lithologic discrimination (Margrave et al., 2001).

Using a simplified model and incorporating all available data might not be enough to overcome the nonuniqueness problem and produce a reasonable answer. In most cases regularisation is also necessary. Regularising the inverse problem means finding a physical meaningful stable solution (Tenorio, 2001). A different approach to handle nonuniqueness is Bayesian formulation and uncertainty estimation. In the Bayesian statistics the model parameters and AVA data will no longer be treated as deterministic constants but will have statistical distributions assigned to them. The end result will not only consist of an estimate of the model parameters but also an estimate of the uncertainty in the parameters. It is very important to note that a Bayesian approach does not remove the nonuniqueness, it only helps identifying it.

Previous work with inversion of reflection coefficients involves different models and solution algorithms. Smith & Gidlow (1987) used a linear approximation to the PP reflection coefficient and a least square approach to solve the inversion. Stewart (1990) and Veire & Landrø (2006) extended this to joint PP and PS inversion in a least square setting. Margrave et al. (2001) gave a nice introduction to joint inversion and compared the results with only PP inversion. For a tutorial of least square inversion and how to regularise it see Lines & Treitel (1984) and for a more comprehensive treatment of regularisation Tenorio (2001).

Buland & Omre (2003a; 2003b) formulated the PP inversion using linear approximations, in the Bayesian framework. In the latter they not only estimate the elastic parameters but also the wavelet and noise level in the AVA data, together with uncertainties. For details on Bayesian modeling from a geophysics standpoint see Sen & Stoffa (1996) and from a statistical standpoint, Robert & Casella (1999) and Liu (2001).

Our approach to the inversion of reflection coefficients is to use an isotropic, quadratic approximation to the Zoeppritz equations instead of the linear ones used earlier. Stovas & Ursin (2001; 2003) gave implicit, second order expressions for reflection and transmission coefficients in both isotropic and transversely isotropic media. They showed that quadratic approximations are superior to the linear ones for intermediate pre-critical reflection angles, but the number of parameters are still three as for the linear approximations. However, for a comparison, we also perform the inversion using a linear approximation and the exact Zoeppritz equations.

We formulate the inversion in a Bayesian framework following Buland & Omre (2003b) and test both PP and joint PP and PS inversion. The use of statistical distributions enable us to impose spatial correlation and correlations between model parameters and reflection angles in a very natural way. The solution algorithm is based on sampling and because of the nonlinear model we use the Metropolis-Hastings algorithm (Robert & Casella, 1999; Liu, 2001) following closely the work of Tjelmeland & Eidsvik (2005). We define our computational domain to be a two dimensional surface, e.g. the top reservoir. The reason is to avoid the wavelet estimation and the convolution in the modeling and instead focus on the

nonlinearity. Our parameters to invert for are  $\mathbf{m}$ ,  $\sigma_e^2$ , and  $\sigma_m^2$ . The first one is the three elastic parameters in the Zoeppritz approximation. The two last parameters are data driven scalars and their main objective is to stabilise the inversion algorithm, but can to some extent quantify the noise level in the reflection coefficients.

In the following sections we first give explicit expressions for quadratic approximations of the reflection coefficients. We then define the Bayesian model and describe the statistical inversion algorithm. In the numerical examples we define four synthetic test cases, show the inversion results, and conclude with a discussion of the results.

## 2 MODEL

The parametrization we are using for the reflection coefficients is in P-wave and S-wave impedance and density, as suggested by Dębski & Tarantola (1995). Stovas & Ursin (2003) derived implicit second order expressions for reflections between two transversely isotropic media. Explicit expressions for PP and PS-reflections simplified for two isotropic media, read

$$\begin{aligned}
r_{\text{PP}} = & \frac{1}{2 \cos^2 \theta_p} \frac{\Delta I_\alpha}{\bar{I}_\alpha} - 4 \sin^2 \theta_s \frac{\Delta I_\beta}{\bar{I}_\beta} - \frac{1}{2} \tan^2 \theta_p (1 - 4\gamma^2 \cos^2 \theta_p) \frac{\Delta \rho}{\bar{\rho}} \\
& + \tan \theta_p \tan \theta_s \left[ 4\gamma^2 (1 - (1 + \gamma^2) \sin^2 \theta_p) \left( \frac{\Delta I_\beta}{\bar{I}_\beta} \right)^2 \right. \\
& \quad - 4\gamma^2 (1 - (\frac{3}{2} + \gamma^2) \sin^2 \theta_p) \left( \frac{\Delta I_\beta}{\bar{I}_\beta} \frac{\Delta \rho}{\bar{\rho}} \right) \\
& \quad \left. + \left( \gamma^2 (1 - (2 + \gamma^2) \sin^2 \theta_p) - \frac{1}{4} \right) \left( \frac{\Delta \rho}{\bar{\rho}} \right)^2 \right]
\end{aligned} \tag{1}$$

$$\begin{aligned}
r_{\text{PS}} = & \sqrt{\tan \theta_p \tan \theta_s} \left\{ \left[ (1 - \cos \theta_s (\cos \theta_s + \gamma \cos \theta_p)) \left( 2 \frac{\Delta I_\beta}{\bar{I}_\beta} - \frac{\Delta \rho}{\bar{\rho}} \right) - \frac{1}{2} \frac{\Delta \rho}{\bar{\rho}} \right] \right. \\
& + \frac{1}{2} \left[ (1 - \cos \theta_s (\cos \theta_s - \gamma \cos \theta_p)) \left( 2 \frac{\Delta I_\beta}{\bar{I}_\beta} - \frac{\Delta \rho}{\bar{\rho}} \right) - \frac{1}{2} \frac{\Delta \rho}{\bar{\rho}} \right] \\
& \times \left[ \frac{1}{2 \cos^2 \theta_p} \frac{\Delta I_\alpha}{\bar{I}_\alpha} + \left( \frac{1}{2 \cos^2 \theta_s} - 8 \sin^2 \theta_s \right) \frac{\Delta I_\beta}{\bar{I}_\beta} \right. \\
& \quad \left. \left. + \left( 4 \sin^2 \theta_s - \frac{1}{2} (\tan^2 \theta_p + \tan^2 \theta_s) \right) \frac{\Delta \rho}{\bar{\rho}} \right] \right\}
\end{aligned} \tag{2}$$

where  $\gamma = \bar{\beta}/\bar{\alpha}$  is the background  $v_S/v_P$ -ratio,  $\theta_p$  is the angle of the incoming P-wave (and also the reflected P-wave because of isotropic medium) and  $\theta_s$  is the angle of the reflected S-wave.  $I_\alpha$  is P-wave impedance,  $I_\beta$  is S-wave impedance and  $\rho$  is density.  $\Delta I_\alpha$  is difference between the lower and upper medium and  $\bar{I}_\alpha$  denotes the average, equal definitions for  $I_\beta$  and  $\rho$ . Appendix A shows how to derive the of reflection coefficients between two isotropic, elastic media for other parametrizations.

In (1) and (2) the normalization is with respect to vertical energy flux. The relation to the more common amplitude normalized expressions is only a scalar factor. For  $r_{PP}$  this constant is equal to one and the linear terms, after a changes of parameters, are equal to the amplitude normalized ones found in Aki & Richards (1980).

The variables to invert for are denoted  $\mathbf{m}$ ,  $\sigma_e^2$ , and  $\sigma_m^2$ . The first one is the obvious and consists of parameters found in (1) and (2),  $m = \{\frac{\Delta I_\alpha}{I_\alpha}, \frac{\Delta I_\beta}{I_\beta}, \frac{\Delta \rho}{\rho}\}$ . It is defined on a two dimensional spatial grid with dimensions  $n_y \times n_x$ ,

$$\mathbf{m} = \{m_{ij} \in \mathbb{R}^{D_m}; \quad i = 1..n_y, j = 1..n_x\}. \quad (3)$$

$D_m$  is the number of unknown in each node and equal to three for the isotropic approximations (1) and (2). The two last parameters,  $\sigma_e^2$  and  $\sigma_m^2$ , are scalars quantifying variance levels. In our Bayesian formulation two covariance matrices have to be specified, and the choice made will influence the solution. We have therefore assumed these two matrices to be known only up to a multiplicative variance factor and let the estimation of the two factors be a part of the inversion procedure. In theory, the full covariance matrices could also be estimated this way but we have chosen to do this only for the variance levels, denoted  $\sigma_e^2$  for the data errors and  $\sigma_m^2$  for the parameter errors.

The input to the inversion, the measured data, is denoted  $\mathbf{d}$ . It is defined over the same spatial grid as for  $\mathbf{m}$ ,

$$\mathbf{d} = \{d_{ij} \in \mathbb{R}^{D_d}; \quad i = 1..n_y, j = 1..n_x\}. \quad (4)$$

$D_d$  is the sum measured PP and PS angles,  $d = \{r_{PP}(\theta), r_{PS}(\theta)\}$  in each node. It is important

that we have more measurements than parameters, i.e.  $D_d \geq D_m$ , otherwise the inversion will be underdetermined.

The forward model is the link from  $\mathbf{m}$  to  $\mathbf{d}$ . In the most exact case with full Zoeppritz it is written

$$\mathbf{d} = f_z(\mathbf{m}) + \mathbf{e}_z, \quad (5)$$

where  $f_z$  is the Zoeppritz equations. The term  $\mathbf{e}_z$  is the observation errors related to the always present noise in measurements. In theory it will also consist of modelling errors but these are neglected under the assumption that the approximations in Zoeppritz are valid.

Our focus will be on the quadratic approximations and the corresponding forward model is

$$\mathbf{d} = f(\mathbf{m}) + \mathbf{e}, \quad (6)$$

where  $f$  is (1) and (2) in the case of both PP and PS reflections. The error term can be approximated with  $\mathbf{e} \approx \mathbf{e}_z$  as long as the quadratic approximations are good. We can also formulate the linear forward model

$$\mathbf{d} = F\mathbf{m} + \mathbf{e}', \quad (7)$$

where  $F$  is a matrix containing only the linear part of (1) and (2). In general,  $\mathbf{e}'$  is not approximating  $\mathbf{e}_z$  as good as  $\mathbf{e}$  does.

A common way of inverting for  $\mathbf{m}$  is to minimise  $\mathbf{e}$  in (6) in an iterative nonlinear least squares algorithm, but this will not produce uncertainty measures. Therefore, we will reformulate the problem in a Bayesian framework. In such a formulation the parameters and measured data are no longer treated as being fixed, but have statistical distributions assigned to them. These distributions can be divided into three groups: prior, likelihood, and posterior.

Prior distributions reflect the information known before the data is measured. This information can be theoretically founded, e.g. velocities can not be negative, or case dependent, e.g. information from the geological setting. The purpose is to restrict the search space to

only reasonable values. However, it has also a down side, prior distributions can greatly influence the final solution. If the prior distribution is chosen to have very small variance, the solution will be close to it, regardless of the information in the measured data. The parameter  $\sigma_m^2$  is introduced in the Bayesian model to help overcome this problem by adjusting the variance in the prior. In our formulation the prior distribution is denoted  $\pi(\mathbf{m}, \sigma_e^2, \sigma_m^2)$ , where  $\pi$  represent any distribution.

Likelihood distributions are the second group. This is the analog to the forward model. In our Bayesian setting it is expressed  $\pi(\mathbf{d}|\mathbf{m}, \sigma_e^2)$  and describes the distribution of  $\mathbf{d}$  given the parameters  $\mathbf{m}$  and  $\sigma_e^2$ . The parameter  $\sigma_e^2$  is introduced to the likelihood for the same reason as  $\sigma_m^2$  was introduced to the prior.

The last distribution is the posterior and is the distribution we would like assess in the inversion procedure. It is the reverse of the likelihood, i.e. given  $\mathbf{d}$ , what is the distribution of  $\mathbf{m}$ ,  $\sigma_e^2$ , and  $\sigma_m^2$ , and is expressed  $\pi(\mathbf{m}, \sigma_e^2, \sigma_m^2|\mathbf{d})$ . Relating the posterior to the prior and likelihood is done by applying the Bayes rule to the posterior distribution, and the final answer is

$$\pi(\mathbf{m}, \sigma_e^2, \sigma_m^2|\mathbf{d}) \propto \pi(\mathbf{d}|\mathbf{m}, \sigma_e^2) \pi(\mathbf{m}, \sigma_e^2, \sigma_m^2). \quad (8)$$

The factor  $\pi(\mathbf{d})$  becomes an unknown constant, that we skip, and hence the proportionality in (8). It can be simplified by assuming some of the parameters to be independent. The details of the Bayesian formulation and choices of distributions are found in Appendix B.

### 3 INVERSION ALGORITHM

Because of the nonlinearity in the likelihood, no analytical solution to the posterior is known. To approximate it we have to produce samples based on the prior and likelihood information, see equation (8), to build the posterior distribution. It is sampled with a Metropolis-Hastings algorithm (Robert & Casella, 1999; Liu, 2001) which sequentially updates  $\mathbf{m}$ ,  $\sigma_e^2$ , and  $\sigma_m^2$ , keeping the two others fixed when updating.

A Metropolis-Hastings algorithm consists of two steps:

- (i) Propose a new sample
- (ii) Accept the sample with a probability  $p$

In the first step essentially any distribution can be used, the acceptance probability in the second step will correct for the errors introduced in the proposal distribution. If one proposes new values from the full conditionals (the distribution of parameter to propose given all other parameters) all proposals are accepted. There is, however, several reasons why this proposal distribution is not always used. It can be nonlinear and impossible to generate sample from, or very complicated or high dimensional and thereby very expensive to sample from.

Back to our problem, we will sample the posterior by sequentially update  $\mathbf{m}$ ,  $\sigma_e^2$ , and  $\sigma_m^2$ , keeping the two others fixed. That is, instead of sampling  $\pi(\mathbf{m}, \sigma_e^2, \sigma_m^2 | \mathbf{d})$  we will sample  $\pi(\mathbf{m} | \mathbf{d}, \sigma_e^2, \sigma_m^2)$ ,  $\pi(\sigma_e^2 | \mathbf{d}, \mathbf{m}, \sigma_m^2)$ , and  $\pi(\sigma_m^2 | \mathbf{d}, \mathbf{m}, \sigma_e^2)$  iteratively. After an update of all three we will have new sample from the posterior (8).

In the process of sampling the three distributions we again use the Metropolis-Hastings algorithm. When sampling  $\mathbf{m}$  we can not use the full conditional since the likelihood contains the nonlinear forward model. We therefore have to use a different distribution and the natural choice is a linearised posterior distribution involving the linear forward model (7) in the likelihood. This also results in an acceptance probability less than 1.

For the two last posterior distributions the situation is different. Here we are able to produce samples from  $\pi(\sigma_e^2 | \mathbf{d}, \mathbf{m}, \sigma_m^2)$  and  $\pi(\sigma_m^2 | \mathbf{d}, \mathbf{m}, \sigma_e^2)$ . The sampling of  $\sigma_e^2$  and  $\sigma_m^2$  are therefore examples of Gibbs updates.

Details about the sampling procedure are found in Appendix C and the linear proposal distribution for  $\mathbf{m}$  is derived in Appendix D.

## 4 NUMERICAL MODEL

We will use a synthetic model to test the inversion algorithm. From a chosen true  $\mathbf{m}$  we can use the exact Zoeppritz equations to generate synthetic measurements  $\mathbf{d}$ . Since the isotropic Zoeppritz equations have 5 parameters we have to fix two variables in addition to the three in  $\mathbf{m}$ . We have chosen the P-wave velocity in the upper medium and the background  $v_P/v_S$



ratio to be constant. A total of four cases will be generated, that is only PP data and both PP and PS data, each of them with and without noise added. For all these four cases we will compare inversions with three different models: linear, quadratic, and exact Zoeppritz.

The truth we have chosen is, as discussed earlier, parametrized in contrasts in P-wave and S-wave impedance and density and they are all ranging from 0.2 to 0.5 as shown in Fig. 1. These contrasts are very strong and the purpose is to test the nonlinear inversion algorithm. The size of the computational grid is  $n_x = n_y = 100$  with a spacing of 25m in each direction.

In the Metropolis-Hastings algorithm the linear model is used to propose an update of the posterior, and it is accepted with a certain probability. To control this probability, and hence the acceptance rate, is very important. If it is equal to 1 all the proposals will be accepted and the result is linear inversion. On the contrary, if it is close to 0 it will produce very few updates which leads to poor convergence and run time problems. The main reason for introducing the two scalars  $\sigma_e^2$  and  $\sigma_m^2$  and their distributions is to help overcome this problem. They are both data driven and will help control the acceptance rate and stabilise the inversion algorithm. To some extent  $\sigma_e^2$  will also quantify the noise level, but it is always with respect to the forward model used. More specifically, it quantifies how well the model can reproduce  $\mathbf{d}$  from the posterior  $\mathbf{m}$ , but this does not necessary quantify the bias.

The expected value of the prior of  $\mathbf{m}$  in all cases is half of the true  $\mathbf{m}$  in Fig. 1. The reason for this is to see if the inversion relies too heavily on prior information, if this is the case it will show as a large bias.

Our first set of measurements is PP reflections without any noise. Fig. 2 shows the reflection coefficients and the corresponding bias in both the linear and quadratic approximations. We use four angles from  $0^\circ$  to  $55^\circ$  and the reason for skipping the bias plots for  $\theta = 0^\circ$  is that with P-wave impedance as one of the parameters both approximations are exact at normal incidence. For the three other angles we see that the quadratic approximations are better than the linear, but when critical angle is approached, in the two lower corners, even the quadratic will have large bias.

The second example we will test is PP inversion when noise is included. In the prior of  $\mathbf{m}$  we add spatially correlated normal distributed noise to break the smoothness. The noise added to the measurements  $\mathbf{d}$  is normally distributed and here it is both spatially correlated and correlated between the angles. Fig. 3 shows  $\mathbf{d}$  when the noise is included. The variance is constant but is most visible for the normal incidence reflection where the range of values is smallest.

After inverting PP reflections we will turn to joint PP and PS inversion. We will use the four coefficients in Fig. 2 and include three PS reflections with incoming P-waves between  $20^\circ$  and  $55^\circ$ . The idea is to see what additional information the PS reflections will bring. Fig. 4 displays the exact Zoeppritz reflection coefficients together with the bias in the approximations, similar to the PP case. Here it is even more clear how superior the quadratic approximations are.

The last synthetic example is joint inversion of reflection coefficients including noise. The prior used is the same as in the previous case with noise. Also for the measurements the situations is almost equal. The variance is still constant, the noise is spatially correlated and we have correlations between the angles, but the correlations between P-wave and S-wave reflections are set to be zero. Fig. 5 shows the measurements generated for joint inversion of PP and PS reflections including noise.

## 5 PP INVERSION

The result after the inversion is the posterior distribution of  $\mathbf{m}$ ,  $\sigma_e^2$ , and  $\sigma_m^2$ . The two last ones are scalars and the distributions are simple to visualise, but this is not the case for  $\mathbf{m}$  since it is multivariate. We therefore calculate the mean of the distribution and subtracts the truth to produce the bias. Fig. 6 shows the absolute value of this bias in the case without noise. Second, we have plotted the standard deviation of the distribution in Fig. 7. The two last scalar distributions are displayed i Fig. 8.

Case number two that we tested was PP data including noise. The absolute value of the

bias of the mean is displayed in Fig. 9, the standard deviation is in Fig. 10, and the scalar distributions are in Fig. 11.

## 6 JOINT PP AND PS INVERSION

The two last synthetic examples are joint inversion with and without noise included. As for the two PP cases we display the absolute value of the bias and the standard deviation for  $\mathbf{m}$  and the full distributions for  $\sigma_e^2$  and  $\sigma_m^2$ . Figs 12 and 13 show the absolute value of the bias and the standard deviation in the case without noise and Fig. 14 the corresponding scalar distributions.

Figs 15, 16, and 17 show the same plots for the last case, joint inversion of PP and PS data with noise added.

## 7 DISCUSSION

In this section we will compare the results from all the four synthetic test cases. We will start by looking at  $\mathbf{m}$  in the two cases without noise followed by the two with noise and conclude by looking at  $\sigma_e^2$  and  $\sigma_m^2$ .

The absolute value of the bias in PP inversion and joint inversion, both without noise, found in Figs 6 and 12, have interesting common features and differences. Because of the parametrization in P-wave impedance we have in practice no bias for this parameter, independent of forward model and synthetic case example. For the two other parameters the bias is several orders of magnitude higher. In the PP case the performance of the three forward models are fairly equal for these two parameters, with the nonlinear ones slightly better as expected. In the joint inversion the situation is different. The linear inversion is actually worse, while the nonlinear ones have improved substantially. In the S-wave impedance we see some effects of too low acceptance rate in the areas to the right. Looking at Fig. 4 we see that this coincides with the areas where the linear model has large bias.

Next, we keep these previous results in mind and compare the corresponding standard deviations in Figs 7 and 13. First of all we note that the additional PS information reduces

the standard deviation for all three parameters. Second, the standard deviation for the P-wave impedance are much smaller compared with the two other parameters, as for the bias. For the linear models the deviations are almost constant. This is because we have chosen space invariant noise and prior distributions and the linear likelihood. The reason for not being exactly constant is the sampling based algorithm we are using. In Fig. 7 the linear inversion shows a lower standard deviations for S-wave impedance than for the density. Comparing these two with the corresponding bias we see that this is a misleading result. The uncertainty in the nonlinear inversions yields a far more realistic result. This is an effect of using a linear forward model which is incorrect. The fit to the data is very good, but the parameter values are incorrect (they have large bias).

Turning to the two last synthetic cases, the inversion of data with noise added, we first look at the bias in Figs 9 and 15. One important change is the scale of the P-wave impedance, the bias is now of the same magnitude as for the S-wave impedance and the density, but the parameter that is most difficult to resolve still seems to be the S-wave impedance. This effect is less pronounced in the joint inversion case. When comparing the three models we see that the differences are minor among them. The amount of noise added is such that the model used is less important.

Looking at the standard deviation in Figs 10 and 16 we see the same feature as in the two first cases, the introduction of PS information greatly reduces the uncertainty for all three models and all three parameters. In addition we note that in this case with noise added, P-wave impedance has the lowest uncertainty while S-wave impedance has the highest.

Last we will briefly comment on the two last scalar parameters,  $\sigma_e^2$  and  $\sigma_m^2$ . As mentioned earlier, the main reason for including these are to help stabilising the inversion. They are relative numbers and their value will depend on the prior distributions and forward models used. The reason for displaying them is for completeness since they are a part of the Bayesian model.

## 8 CONCLUSION

In our approach to inversion of reflection coefficients we have reformulated the problem in a Bayesian framework. The use of prior and likelihood distributions enables us to assess the uncertainty in the inversion result, the posterior. At the same time it enables us easily to impose correlations and covariances in the modeling. We have also used new quadratic approximations to the Zoeppritz equations and compared them with both the linearised and exact equations. Because of the nonlinear forward models we have tested the performance of the Metropolis-Hastings algorithm to sample from the nonlinear likelihood. Last, we have also included inversion of joint PP and PS and compared with the more common PP inversion.

Our work shows that the Metropolis-Hastings algorithm works for this type of application, but it is crucial to control the acceptance rate to achieve proper convergence. The quadratic approximations outperformed the linear ones when the reflection data have a low noise level, otherwise the inversion will yield more or less equal results. On the other hand, the differences between exact Zoeppritz equations and the quadratic approximations are minor, even in the case without noise. Last, by including PS reflection data in the inversion the bias and the uncertainties are greatly reduced.

## ACKNOWLEDGMENT

We wish to thank BP, Hydro, Schlumberger, Statoil, and The Research Council of Norway for their support through the Uncertainty in Reservoir Evaluation (URE) project.

## REFERENCES

- Aki, K. & Richards, P. G., 1980. *Quantitative Seismology: Theory and Methods*, W. H. Freeman and Co.
- Buland, A. & Omre, H., 2003a. Bayesian linearized AVO inversion, *Geophysics*, **68**, 185–198.
- Buland, A. & Omre, H., 2003b. Joint AVO inversion, wavelet estimation and noise-level estimation using a spatially coupled hierarchical bayesian model, *Geophysical Prospecting*, **51**, 531–550.

- Dębski, W. & Tarantola, A., 1995. Information on elastic parameters obtained from the amplitudes of reflected waves, *Geophysics*, **60**, 1426–1436.
- Hampson, D., 1991. AVO inversion, theory and practice, *The Leading Edge*.
- Lines, L. R. & Treitel, S., 1984. A review of least-squares inversion and its application to geophysical problems, *Geophysical Prospecting*, **32**, 159–186.
- Liu, J. S., 2001. *Monte Carlo strategies in scientific computing*, Springer.
- Margrave, G. F., Stewart, R. R., & Larsen, J. A., 2001. Joint PP and PS seismic inversion, *The Leading Edge*.
- Robert, C. P. & Casella, G., 1999. *Monte Carlo statistical methods*, Springer.
- Sen, M. K. & Stoffa, P. L., 1996. Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion, *Geophysical Prospecting*, **44**, 313–350.
- Smith, G. C. & Gidlow, P. M., 1987. Weighted stacking for rock property estimation and detection of gas, *Geophysical Prospecting*, **35**, 993–1014.
- Stewart, R. R., 1990. Joint P and P-SV seismic inversion, Tech. rep., CREWES.
- Stovas, A. & Ursin, B., 2001. Second-order approximations of the reflection and transmission coefficients between two visco-elastic isotropic media, *Journal of Seismic Exploration*, **9**, 223–233.
- Stovas, A. & Ursin, B., 2003. Reflection and transmission responses of layered transversely isotropic viscoelastic media, *Geophysical Prospecting*, **51**, 447–477.
- Tenorio, L., 2001. Statistical regularization of inverse problems, *SIAM Review*, **43**, 347–366.
- Tjelmeland, H. & Eidsvik, J., 2005. Directional Metropolis-Hastings update for posteriors with non-linear likelihoods, in *Geostatistics Banff 2004*, edited by O. Leuangthong & C. V. Deutsch.
- Ursin, B. & Tjåland, E., 1996. The information content of the elastic reflection matrix, *Geophysical Journal International*, **125**, 214–228.
- Veire, H. H. & Landrø, M., 2006. Simultaneous inversion of PP and PS seismic data, *Geophysics*, **71**, R1–R10.

**APPENDIX A: SECOND ORDER REFLECTION COEFFICIENTS**

The complete set of reflection coefficients between two transversely isotropic media for a downgoing wave, derived by Stovas & Ursin (2001), reads

$$\mathbf{R}_D = \begin{bmatrix} r_{PP} & r_{PS} \\ r_{SP} & r_{SS} \end{bmatrix} \approx - \begin{bmatrix} g_{11} - g_{12}f & g_{12} + \frac{1}{2}f(g_{11} - g_{22}) \\ g_{12} + \frac{1}{2}f(g_{11} - g_{22}) & g_{22} + g_{12}f \end{bmatrix}. \quad (\text{A1})$$

The parametrization we have chosen to express  $f$ ,  $g_{11}$ ,  $g_{22}$ , and  $g_{12}$  in is shear and plane wave modulus together with density. This choice results in the most compact expressions when simplified to isotropic media:

$$\begin{bmatrix} g_{11} \\ g_{22} \\ g_{12} \\ f \end{bmatrix} = \begin{bmatrix} \frac{1}{4\cos^2\theta_p} & -2\sin^2\theta_s & \frac{1}{4}(1 - \tan^2\theta_p) \\ 0 & 2\sin^2\theta_s - \frac{1}{4\cos^2\theta_s} & -\frac{1}{4}(1 - \tan^2\theta_s) \\ 0 & k(\cos\theta_s(\cos\theta_s + \gamma\cos\theta_p) - 1) & \frac{1}{2}k \\ 0 & k(\cos\theta_s(\cos\theta_s - \gamma\cos\theta_p) - 1) & \frac{1}{2}k \end{bmatrix} \begin{bmatrix} \frac{\Delta M}{\bar{M}} \\ \frac{\Delta\mu}{\bar{\mu}} \\ \frac{\Delta\rho}{\bar{\rho}} \end{bmatrix} \quad (\text{A2})$$

with the plane wave modulus,  $M = \rho\alpha^2$ ; the shear modulus,  $\mu = \rho\beta^2$ ; and density,  $\rho$ . The difference is  $\Delta M = M_2 - M_1 = \rho_2\alpha_2^2 - \rho_1\alpha_1^2$  and the average is  $\bar{M} = (\rho_2\alpha_2^2 + \rho_1\alpha_1^2)/2$  with equal definitions for  $\mu$  and  $\rho$ . The background  $v_S/v_P$  ratio is  $\gamma = \bar{\beta}/\bar{\alpha}$  and the constant  $k$  is  $\sqrt{\tan\theta_p \tan\theta_s}$ . See Fig. A1 for the definition of the angles and parameters.

For a downgoing P-wave reflected to P and S-waves the explicit expressions are

$$r_{PP} = \frac{1}{4\cos^2\theta_p} \frac{\Delta M}{\bar{M}} - 2\sin^2\theta_s \frac{\Delta\mu}{\bar{\mu}} + \frac{1}{4}(1 - \tan^2\theta_p) \frac{\Delta\rho}{\bar{\rho}} + \tan\theta_p \tan\theta_s \left[ \gamma^2(1 - (1 + \gamma^2)\sin^2\theta_p) \left( \frac{\Delta\mu}{\bar{\mu}} \right)^2 - \frac{1}{4} \left( \frac{\Delta\rho}{\bar{\rho}} \right)^2 + \sin^2\theta_s \left( \frac{\Delta\mu}{\bar{\mu}} \frac{\Delta\rho}{\bar{\rho}} \right) \right] \quad (\text{A3})$$

and

$$\begin{aligned}
r_{\text{PS}} = & \sqrt{\tan \theta_p \tan \theta_s} \left\{ \left[ (1 - \cos \theta_s (\cos \theta_s + \gamma \cos \theta_p)) \frac{\Delta \mu}{\bar{\mu}} - \frac{1}{2} \frac{\Delta \rho}{\bar{\rho}} \right] \right. \\
& + \frac{1}{2} \left[ (1 - \cos \theta_s (\cos \theta_s - \gamma \cos \theta_p)) \frac{\Delta \mu}{\bar{\mu}} - \frac{1}{2} \frac{\Delta \rho}{\bar{\rho}} \right] \\
& \left. \times \left[ \frac{1}{4 \cos^2 \theta_p} \frac{\Delta M}{\bar{M}} + \left( \frac{1}{4 \cos^2 \theta_s} - 4 \sin^2 \theta_s \right) \frac{\Delta \mu}{\bar{\mu}} + \frac{1}{4} \left( 2 - \tan^2 \theta_p - \tan^2 \theta_s \right) \frac{\Delta \rho}{\bar{\rho}} \right] \right\}.
\end{aligned} \tag{A4}$$

To express in other parametrizations, e.g. (1) and (2), the following relations can be used

$$\begin{aligned}
\frac{\Delta M}{\bar{M}} &= (1 - \frac{4}{3} \gamma^2) \frac{\Delta K}{\bar{K}} + \frac{4}{3} \gamma^2 \frac{\Delta \mu}{\bar{\mu}} \\
\frac{\Delta M}{\bar{M}} &= (1 - 2\gamma^2) \frac{\Delta \lambda}{\bar{\lambda}} + 2\gamma^2 \frac{\Delta \mu}{\bar{\mu}} \\
\frac{\Delta M}{\bar{M}} &= \frac{2(1 - \gamma^2)^2}{\gamma^2} \Delta \sigma + \frac{\Delta \mu}{\bar{\mu}} \\
\frac{\Delta M}{\bar{M}} &= 2 \frac{\Delta \alpha}{\bar{\alpha}} + \frac{\Delta \rho}{\bar{\rho}} \\
\frac{\Delta M}{\bar{M}} &= 2 \frac{\Delta I_\alpha}{\bar{I}_\alpha} - \frac{\Delta \rho}{\bar{\rho}} \\
\frac{\Delta \mu}{\bar{\mu}} &= 2 \frac{\Delta \beta}{\bar{\beta}} + \frac{\Delta \rho}{\bar{\rho}} \\
\frac{\Delta \mu}{\bar{\mu}} &= 2 \frac{\Delta I_\beta}{\bar{I}_\beta} - \frac{\Delta \rho}{\bar{\rho}},
\end{aligned} \tag{A5}$$

where  $K$  is bulk modulus,  $M = K + \frac{4}{3}\mu$ ;  $\lambda$  is Lamé constant,  $K = \lambda + 2\mu$ ;  $\sigma$  is Poisson's ratio,  $\sigma = \frac{1}{2}(M - 2\mu)/(M - \mu)$ ; and  $I$  is impedance,  $I_\alpha = \rho\alpha$  and  $I_\beta = \rho\beta$ .

## APPENDIX B: BAYESIAN FORMULATION

The joint posterior distribution (8) can be simplified to

$$\pi(\mathbf{m}, \sigma_e^2, \sigma_m^2 | \mathbf{d}) \propto \pi(\mathbf{d} | \mathbf{m}, \sigma_e^2) \pi(\mathbf{m} | \sigma_m^2) \pi(\sigma_e^2) \pi(\sigma_m^2), \tag{B1}$$

by assuming the prior of  $\mathbf{m}$  and likelihood of  $\mathbf{d}$  to be independent of  $\sigma_e^2$  and  $\sigma_m^2$  respectively and  $\sigma_e^2$  and  $\sigma_m^2$  to be a priori independent.

We are assuming the error to be multivariate Gaussian distributed with zero mean and



covariance  $\sigma_e^2 \Sigma_e$ ,

$$\pi(\mathbf{e}|\sigma_e^2) = \mathcal{N}(\mathbf{e}; 0, \sigma_e^2 \Sigma_e). \quad (\text{B2})$$

The prior of  $\mathbf{m}$  is also assumed to be a multivariate Gaussian distribution and the fact that the forward model is deterministic together with (B2) makes also the likelihood of  $\mathbf{d}$  multivariate Gaussian

$$\pi(\mathbf{m}|\sigma_m^2) = \mathcal{N}(\mathbf{m}; \boldsymbol{\mu}_m, \sigma_m^2 \Sigma_m) \quad (\text{B3})$$

$$\pi(\mathbf{d}|\mathbf{m}, \sigma_e^2) = \mathcal{N}(\mathbf{d}; f(\mathbf{m}), \sigma_e^2 \Sigma_e).$$

Here  $f(\mathbf{m})$  is the nonlinear forward model in (6),  $\boldsymbol{\mu}_m$  is the prior expected value of  $\mathbf{m}$ , and  $\Sigma_e$  and  $\Sigma_m$  the covariance matrices excluding the level factors  $\sigma_e^2$  and  $\sigma_m^2$ . The  $\Sigma$ 's allow for specifying a priori spatial correlations, correlations between parameters within a grid point, and correlated noise for neighbouring reflection angles, while the level factors,  $\sigma_m^2$  and  $\sigma_e^2$ , will be explored as a part of the inversion. For details about the multivariate Gaussian distribution see Appendix E. A graphical representation of the prior of  $\mathbf{m}$ , the likelihood, and the deterministic relation is shown in the graph in Fig. B1.

The prior of  $\sigma_e^2$  and  $\sigma_m^2$  are assigned inverse gamma distributions

$$\pi(\sigma_e^2) = \mathcal{IG}(\sigma_e^2; \alpha_e, \beta_e) \quad (\text{B4})$$

$$\pi(\sigma_m^2) = \mathcal{IG}(\sigma_m^2; \alpha_m, \beta_m),$$

with corresponding parameters  $\alpha$  and  $\beta$ . For details about the inverse gamma distribution see Appendix E.

## APPENDIX C: SAMPLING

Analytical calculations of the posterior (B1) is not possible. Instead, the joint posterior distributions for  $\mathbf{m}$ ,  $\sigma_e^2$ , and  $\sigma_m^2$  will be sampled iteratively in a Metropolis-Hastings algorithm.

In a simple case of sampling  $\mathbf{x}$  from the posterior  $\pi(\mathbf{x}|\mathbf{y})$ , in our case  $\mathbf{x}$  could be  $\mathbf{m}$ ,  $\sigma_e^2$  or  $\sigma_m^2$  and  $\mathbf{y}$  be  $\mathbf{d}$ , the algorithm consists of the two steps

- (i) Propose a new sample  $\mathbf{x}^u \sim q(\cdot|\mathbf{y})$

(ii) Accept  $\mathbf{x}^u$  with a probability  $p(\mathbf{x}^u|\mathbf{x}) = \min \left\{ 1, \frac{\pi(\mathbf{x}^u|\mathbf{y})}{\pi(\mathbf{x}|\mathbf{y})} \cdot \frac{q(\mathbf{x}|\mathbf{y})}{q(\mathbf{x}^u|\mathbf{y})} \right\}$

where  $q$  is the proposal distribution.

First, since  $\sigma_e^2$  and  $\sigma_m^2$  are assumed to be a posteriori independent they can be sampled in parallel. Their full conditional are

$$\pi(\sigma_e^2|\mathbf{d}, \mathbf{m}) \propto \pi(\mathbf{d}|\mathbf{m}, \sigma_e^2) \pi(\sigma_e^2) \quad (\text{C1})$$

$$\pi(\sigma_m^2|\mathbf{m}) \propto \pi(\mathbf{m}|\sigma_m^2) \pi(\sigma_m^2),$$

and with the choice of likelihood and prior, (B3) and (B4), the posterior also become inverse gamma distributed only with modified parameters. A proof of this is shown in Appendix E.

The explicit posterior expressions read

$$\pi(\sigma_e^2|\mathbf{d}, \mathbf{m}) \sim \mathcal{IG} \left( \sigma_e^2 \left| \alpha + \frac{n_e}{2}, \beta + s_e^2 \frac{n_e}{2} \right. \right) \quad (\text{C2})$$

with

$$s_e^2 = \frac{1}{n_e} (\mathbf{d} - f(\mathbf{m}))^T \Sigma_e^{-1} (\mathbf{d} - f(\mathbf{m})) \quad (\text{C3})$$

and

$$\pi(\sigma_m^2|\mathbf{m}) \sim \mathcal{IG} \left( \sigma_m^2 \left| \alpha + \frac{n_m}{2}, \beta + s_m^2 \frac{n_m}{2} \right. \right) \quad (\text{C4})$$

with

$$s_m^2 = \frac{1}{n_m} (\mathbf{m} - \boldsymbol{\mu}_m)^T \Sigma_m^{-1} (\mathbf{m} - \boldsymbol{\mu}_m) \quad (\text{C5})$$

The scalars are, from (3) and (4),  $n_e = n_x n_y D_d$  and  $n_m = n_x n_y D_m$ . Since the posterior distributions are simply inverse gamma distributed, this will also become our proposal distribution and the acceptance probability will be identical to one. The updates of  $\sigma_e^2$  and  $\sigma_m^2$  are therefore examples of Gibbs updates.

Second, sampling the posterior of  $\mathbf{m}$  is not as simple as for the  $\sigma$ 's. It is not possible to choose the full conditional as the proposal distribution because of the nonlinear likelihood, there is no algorithm to sample from it. The natural choice is then to use the linearised likelihood and the posterior then becomes

$$\pi^{lin}(\mathbf{m}|\mathbf{d}, \sigma_e^2, \sigma_m^2) \propto \pi^{lin}(\mathbf{d}|\mathbf{m}, \sigma_e^2) \pi(\mathbf{m}|\sigma_m^2), \quad (\text{C6})$$

where  $\pi^{lin}(\mathbf{d}|\mathbf{m}, \sigma_e^2)$  is the linearised likelihood using (7).

The use of a simplified proposal distribution also implies that the acceptance rate will be different from zero. In fact, if we try to update the complete  $\mathbf{m}$ , the acceptance rate will be impractically low - if not zero. We therefore have to update only blocks of  $\mathbf{m}$  and repeat this several times to produce an update of  $\mathbf{m}$  in the joint posterior distribution.

More specifically, the proposal distribution is conditioned on  $\mathbf{d}$  and  $\mathbf{m}$  in two regions A and B together with the scalars  $\sigma_e^2$  and  $\sigma_m^2$ ,

$$m_A^u \sim \pi^{lin}(\cdot | m_B, d_A, d_B, \sigma_e^2, \sigma_m^2). \quad (\text{C7})$$

Fig. C1 shows the different blocks. The reason for including B is to impose continuity of the update with respect to the surrounding grid points. Block C does not contribute to the distribution of  $m_A^u$ . For details on the linear update distribution see Appendix D. The acceptance probability reads

$$p(\mathbf{m}^u | \mathbf{m}) = \min \left\{ 1, \frac{\pi(\mathbf{m}^u | \mathbf{d}, \sigma_e^2, \sigma_m^2)}{\pi(\mathbf{m} | \mathbf{d}, \sigma_e^2, \sigma_m^2)} \cdot \frac{\pi^{lin}(m_A | m_B, d_A, d_B, \sigma_e^2, \sigma_m^2)}{\pi^{lin}(m_A^u | m_B, d_A, d_B, \sigma_e^2, \sigma_m^2)} \right\}, \quad (\text{C8})$$

where the first fraction is calculated using the exact, nonlinear version of (C6).

## APPENDIX D: LINEAR PROPOSAL DISTRIBUTION

To sample  $m_A^u$  from the distribution (C7) we need the linear forward model (7). For block A and B it reads

$$d_A = F_A m_A + e_A \quad (\text{D1})$$

$$d_B = F_B m_B + e_B.$$

Partitioning the prior of  $\mathbf{m}$  and the noise model yield

$$\begin{bmatrix} m_A \\ m_B \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix}, \begin{bmatrix} \sigma_m^2 \Sigma_{m,AA} & \sigma_m^2 \Sigma_{m,AB} \\ \sigma_m^2 \Sigma_{m,BA} & \sigma_m^2 \Sigma_{m,BB} \end{bmatrix} \right) \quad (\text{D2})$$

and

$$\begin{bmatrix} e_A \\ e_B \end{bmatrix} \sim \mathcal{N} \left( 0, \begin{bmatrix} \sigma_e^2 \Sigma_{e,AA} & \sigma_e^2 \Sigma_{e,AB} \\ \sigma_e^2 \Sigma_{e,BA} & \sigma_e^2 \Sigma_{e,BB} \end{bmatrix} \right) \quad (\text{D3})$$

We are now able to calculate the mean and correlations between  $m_A$ ,  $m_B$ ,  $d_A$ , and  $d_B$  that we need. For instance

$$\begin{aligned} \mathbb{E}(d_A) &= \mathbb{E}(F_A m_A + e_A) \\ &= F_A \mu_A \end{aligned}$$

$$\begin{aligned} \text{Cov}(m_A, d_B) &= \text{Cov}(m_A, F_B m_B + e_B) \\ &= \sigma_m^2 \Sigma_{m,AB} F_B^T \end{aligned} \tag{D4}$$

$$\begin{aligned} \text{Cov}(d_A, d_B) &= \text{Cov}(F_A m_A + e_A, F_B m_B + e_B) \\ &= \text{Cov}(F_A m_A, F_B m_B) + 2 \times 0 + \text{Cov}(e_A, e_B) \\ &= \sigma_m^2 F_A \Sigma_{m,AB} F_B^T + \sigma_e^2 \Sigma_{e,AB}. \end{aligned}$$

The full distribution becomes

$$\begin{aligned} \begin{bmatrix} m_A \\ m_B \\ d_A \\ d_B \end{bmatrix} &= \mathcal{N} \left( \begin{bmatrix} \mu_A \\ \mu_B \\ F_A \mu_A \\ F_B \mu_B \end{bmatrix}, \begin{bmatrix} \sigma_m^2 \Sigma_{m,AA} \\ \sigma_m^2 \Sigma_{m,BA} \\ \sigma_m^2 F_A \Sigma_{m,AA} \\ \sigma_m^2 F_B \Sigma_{m,BA} \\ \sigma_m^2 \Sigma_{m,AB} & \sigma_m^2 \Sigma_{m,AA} F_A^T & \sigma_m^2 \Sigma_{m,AB} F_B^T \\ \sigma_m^2 \Sigma_{m,BB} & \sigma_m^2 \Sigma_{m,BA} F_A^T & \sigma_m^2 \Sigma_{m,BB} F_B^T \\ \sigma_m^2 F_A \Sigma_{m,AB} & \sigma_m^2 F_A \Sigma_{m,AA} F_A^T + \sigma_e^2 \Sigma_{e,AA} & \sigma_m^2 F_A \Sigma_{m,AB} F_B^T + \sigma_e^2 \Sigma_{e,AB} \\ \sigma_m^2 F_B \Sigma_{m,BB} & \sigma_m^2 F_B \Sigma_{m,BA} F_A^T + \sigma_e^2 \Sigma_{e,BA} & \sigma_m^2 F_B \Sigma_{m,BB} F_B^T + \sigma_e^2 \Sigma_{e,BB} \end{bmatrix} \right), \end{aligned} \tag{D5}$$

which we conveniently group and define by

$$\begin{bmatrix} m_A \\ x_R \end{bmatrix} = \mathcal{N} \left( \begin{bmatrix} \mu_A \\ \mu_R \end{bmatrix}, \begin{bmatrix} \Sigma_{AA} & \Sigma_{AR} \\ \Sigma_{RA} & \Sigma_{RR} \end{bmatrix} \right). \tag{D6}$$

Using the expression for conditional normal distribution, equation (E3), the proposal distribution becomes  $\mathcal{N}(m_A^u; \mu_u, \Sigma_u)$  with

$$\begin{aligned}\mu_u &= \mu_A + \Sigma_{AR}\Sigma_{RR}^{-1}(x_R - \mu_R) \\ \Sigma_u &= \Sigma_{AA} - \Sigma_{AR}\Sigma_{RR}^{-1}\Sigma_{RA}.\end{aligned}\tag{D7}$$

## APPENDIX E: STATISTICAL DISTRIBUTIONS

A multivariate Gaussian variable  $\mathbf{x}$  with expectation vector  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$  has the probability function

$$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\},\tag{E1}$$

where  $n$  is the dimension of  $\mathbf{x}$ .

For two multivariate Gaussian variables  $\mathbf{x}_1$  and  $\mathbf{x}_2$  with joint distribution

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right),\tag{E2}$$

the conditional expectation and variance are

$$\begin{aligned}\boldsymbol{\mu}_{1|2} &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ \boldsymbol{\Sigma}_{1|2} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.\end{aligned}\tag{E3}$$

The inverse gamma probability function is

$$\mathcal{IG}(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{x}\right)^{\alpha+1} e^{-\beta/x}\tag{E4}$$

where  $x \geq 0$ ,  $\alpha > 0$ , and  $\beta > 0$ .

Given the prior distribution  $\sigma^2 \sim \mathcal{IG}(\alpha, \beta)$  and measurements  $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2\boldsymbol{\Sigma})$ , the

posterior distribution of  $\sigma^2$  is

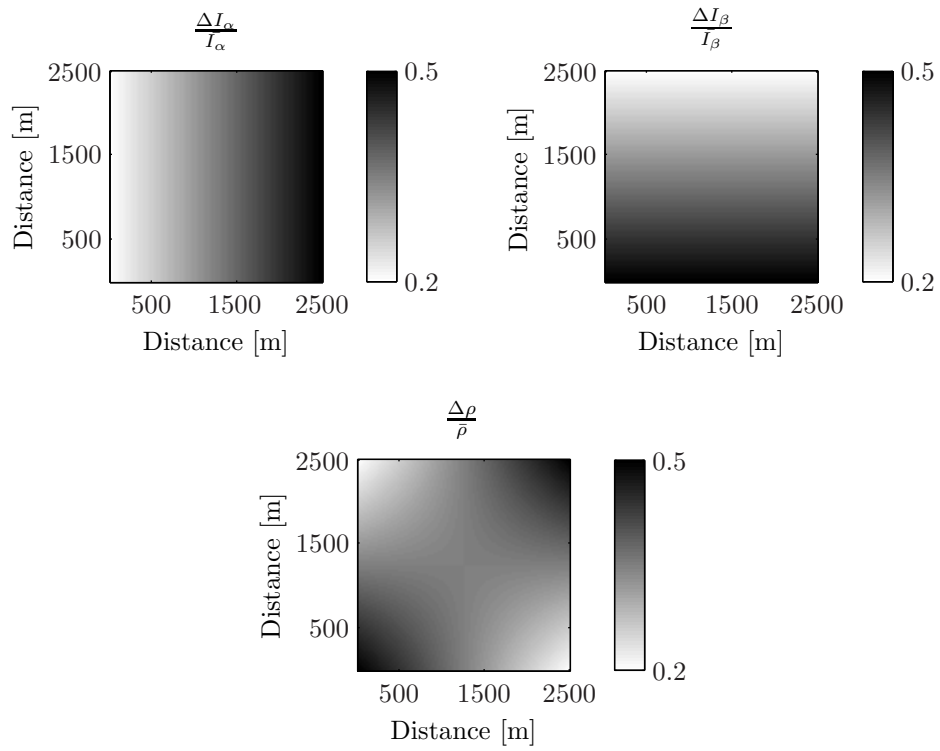
$$\begin{aligned}
\pi(\sigma^2|\mathbf{x}) &\propto \pi(\mathbf{x}|\sigma^2)\pi(\sigma^2) \\
&= \mathcal{N}(\boldsymbol{\mu}, \sigma^2\boldsymbol{\Sigma})\mathcal{IG}(\alpha, \beta) \\
&\propto \frac{1}{(\sigma^2)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}\sigma^{-2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \\
&\quad \times \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp\left\{-\frac{\beta}{\sigma^2}\right\} \\
&\propto \left(\frac{1}{\sigma^2}\right)^{\alpha+1+n/2} \exp\left\{-\sigma^{-2}\left(\beta + s^2\frac{n}{2}\right)\right\} \\
&\propto \mathcal{IG}\left(\sigma^2\left|\alpha + \frac{n}{2}, \beta + s^2\frac{n}{2}\right.\right)
\end{aligned} \tag{E5}$$

where

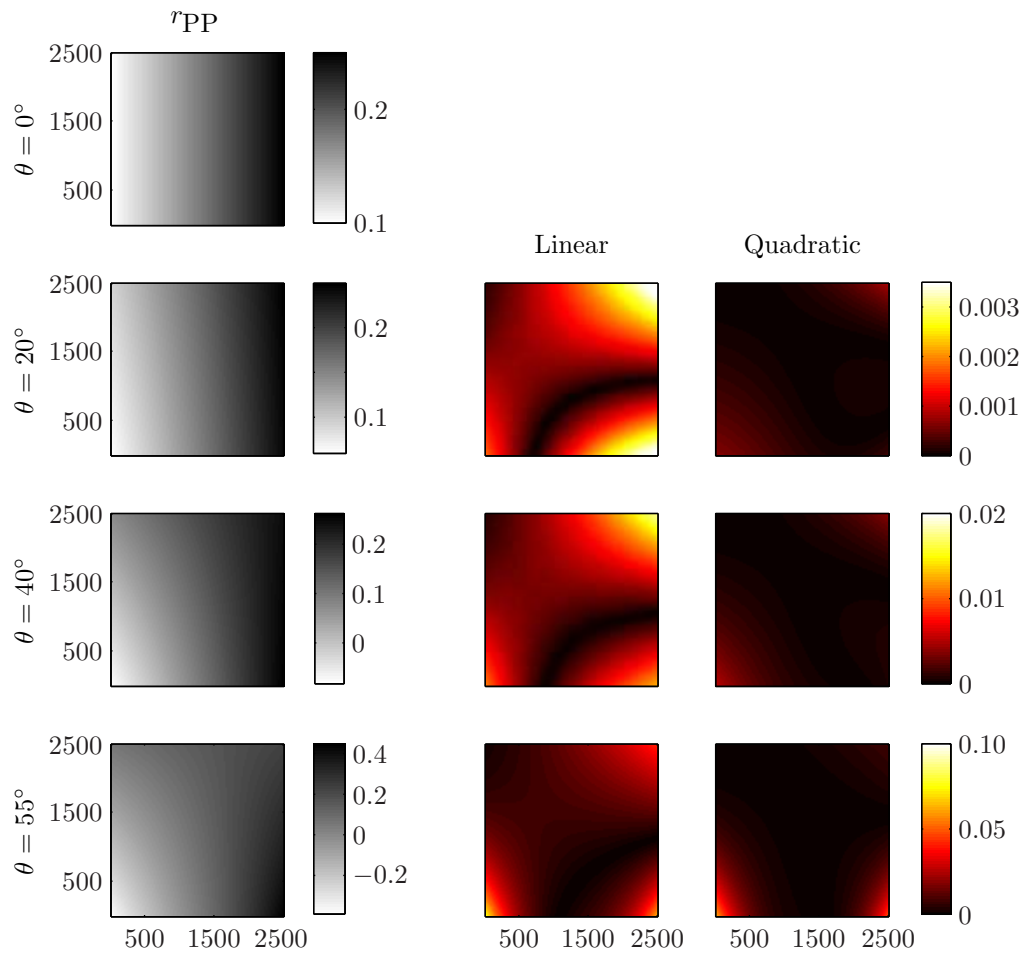
$$s^2 = \frac{1}{n}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \tag{E6}$$

and  $n$  is the dimension of  $\mathbf{x}$ . Clearly, the posterior is also inverse gamma but with modified parameters.

This paper has been produced using the Blackwell Publishing GJI L<sup>A</sup>T<sub>E</sub>X2e class file.

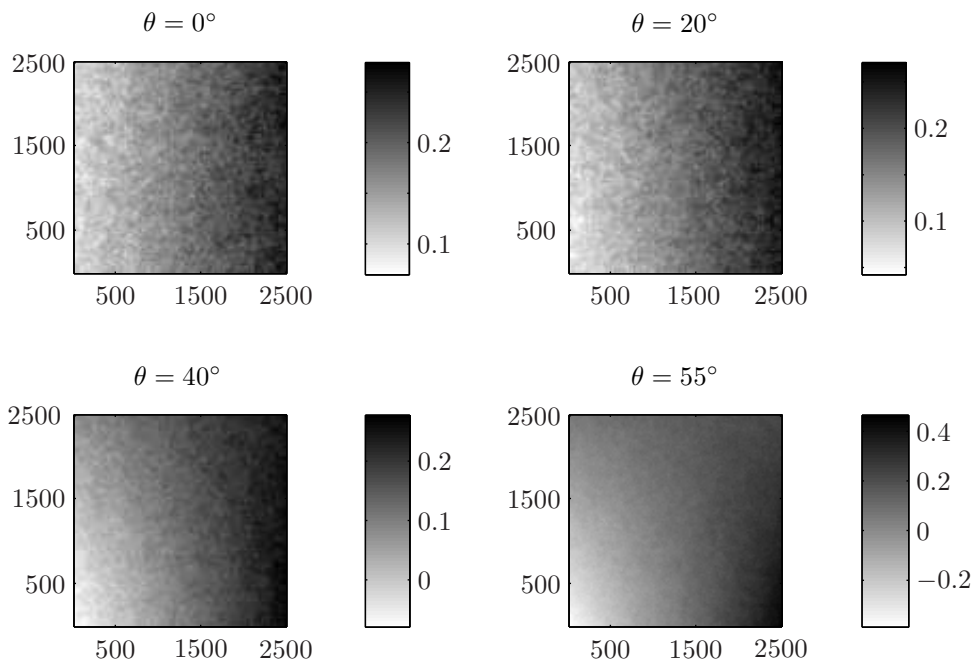


**Figure 1.** The true contrasts in  $\mathbf{m}$ . Upper left is P-wave impedance, upper right is S-wave impedance, and lower is density

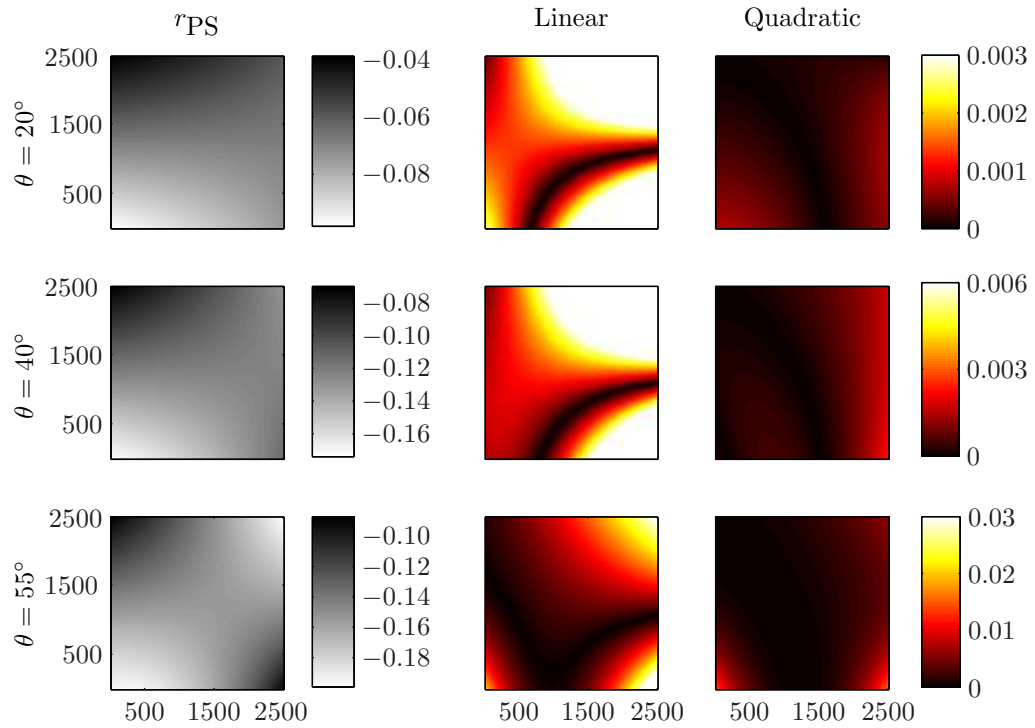


**Figure 2.** To the left is PP reflection coefficients from the Zoeppritz model for 4 different incidence angles. The two right columns show the bias in the linear and quadratic approximations, relative to the Zoeppritz model, for the nonzero angles. For  $\theta = 0^\circ$  the bias is zero.

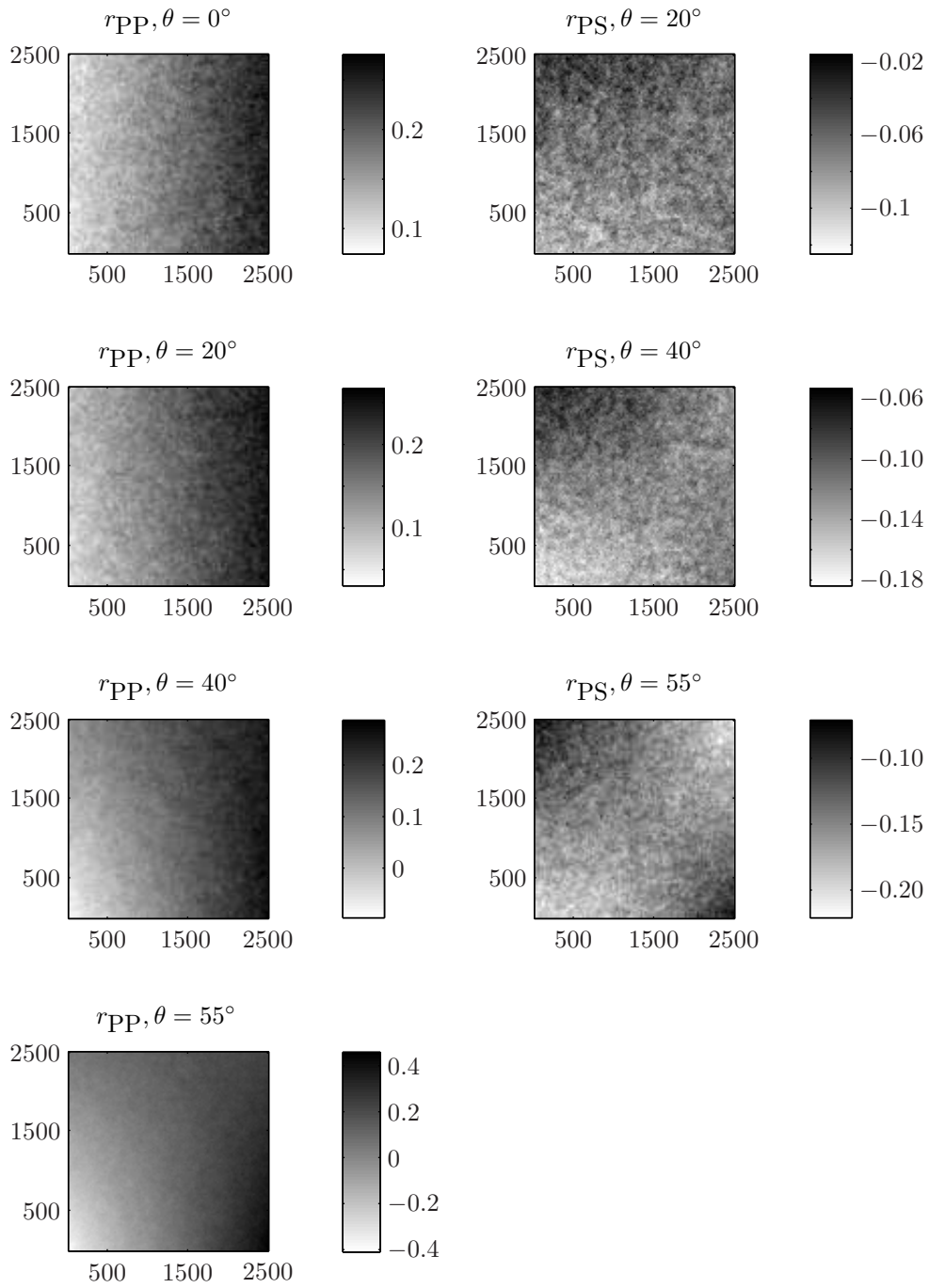




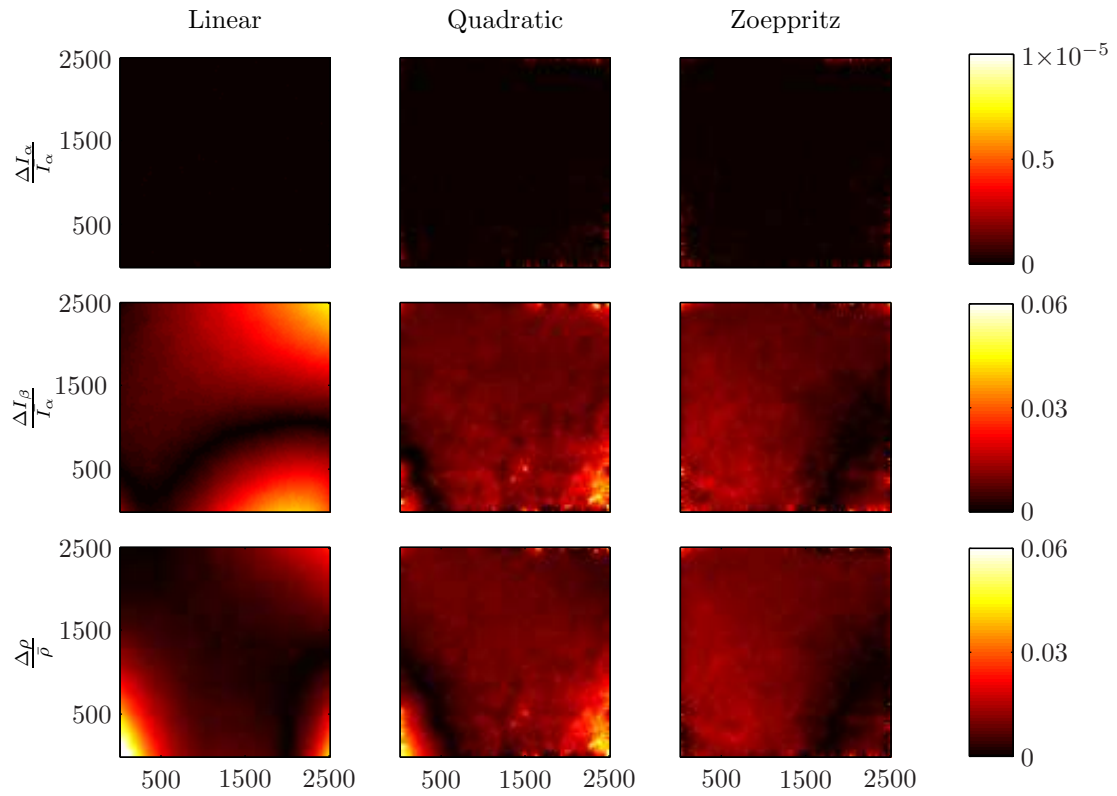
**Figure 3.** PP reflection coefficients from the Zoeppritz model and including multivariate normal distributed noise.



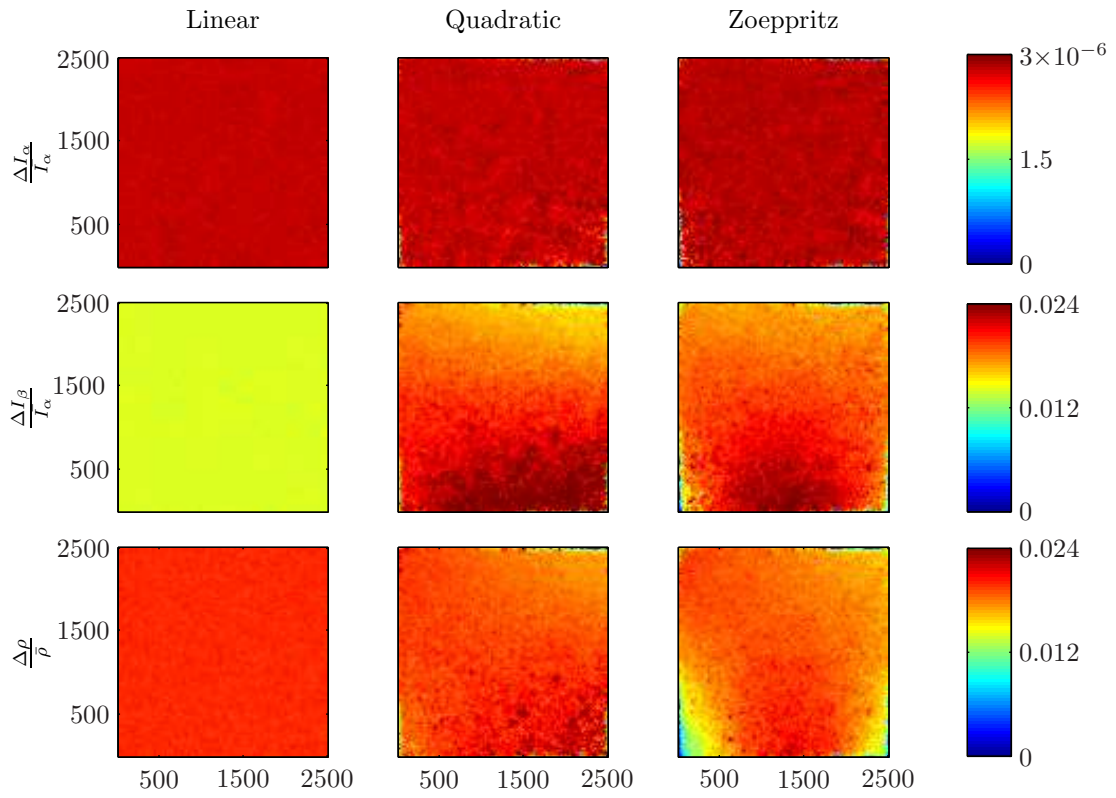
**Figure 4.** To the left is PS reflection coefficients from the Zoeppritz model for 3 nonzero incidence angles. The two right columns show the bias in the linear and quadratic approximations, relative to the Zoeppritz model.



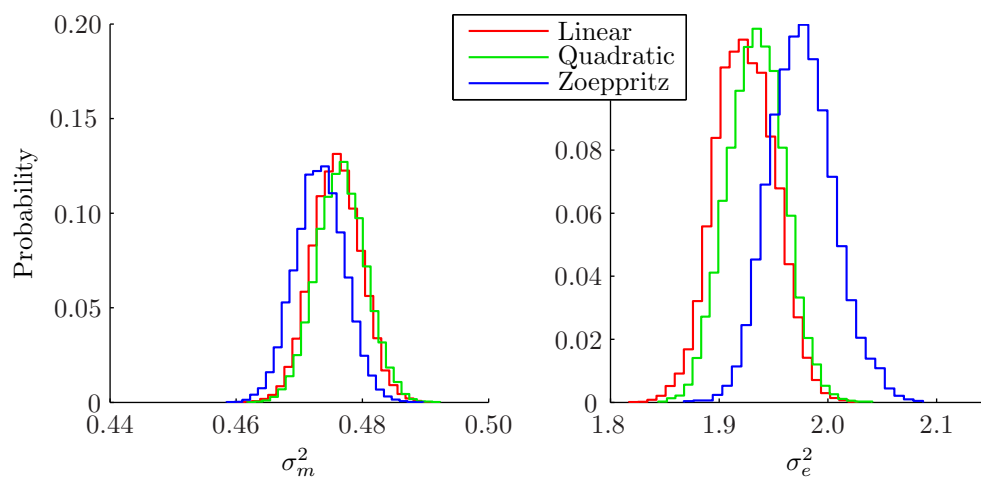
**Figure 5.** PP and PS reflection coefficients from the Zoeppritz model and including multivariate normal distributed noise.



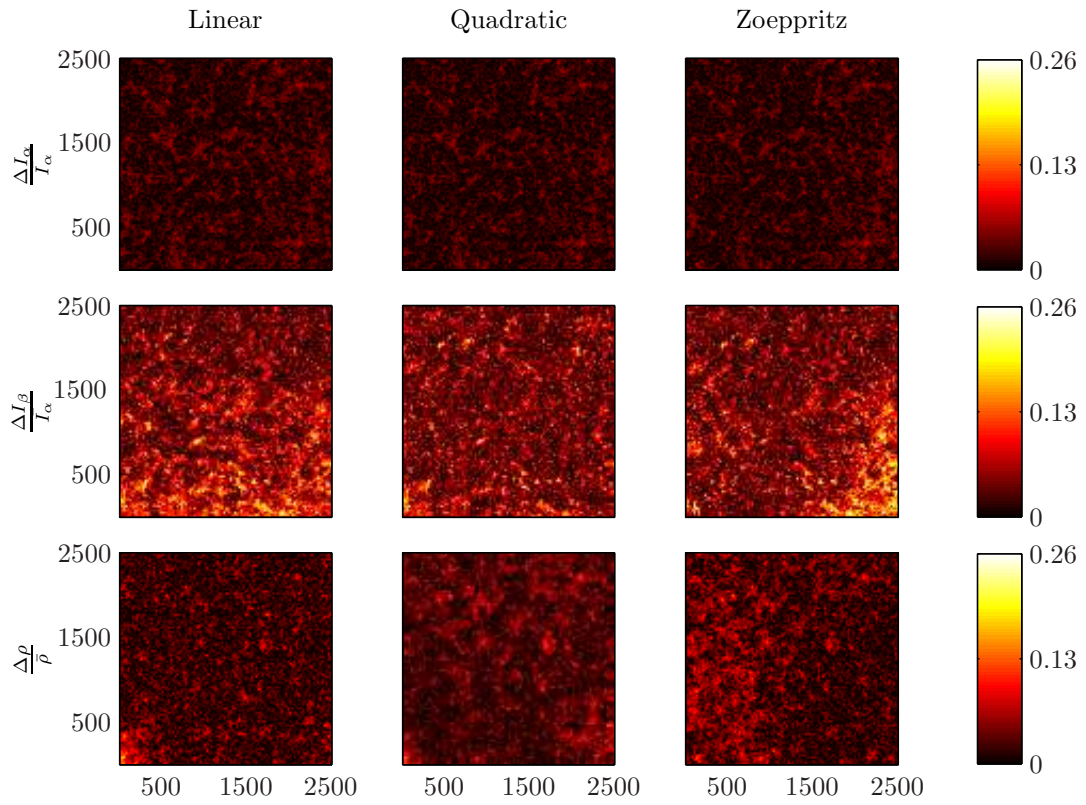
**Figure 6.** Bias in the posterior distribution of  $\mathbf{m}$  from PP inversion without noise. Each column is the result of three inversions using three different models; linear, quadratic, and exact Zoeppritz. The rows displays the three different parameters of  $\mathbf{m}$ ; contrasts in P-wave impedance, S-wave impedance, and density.



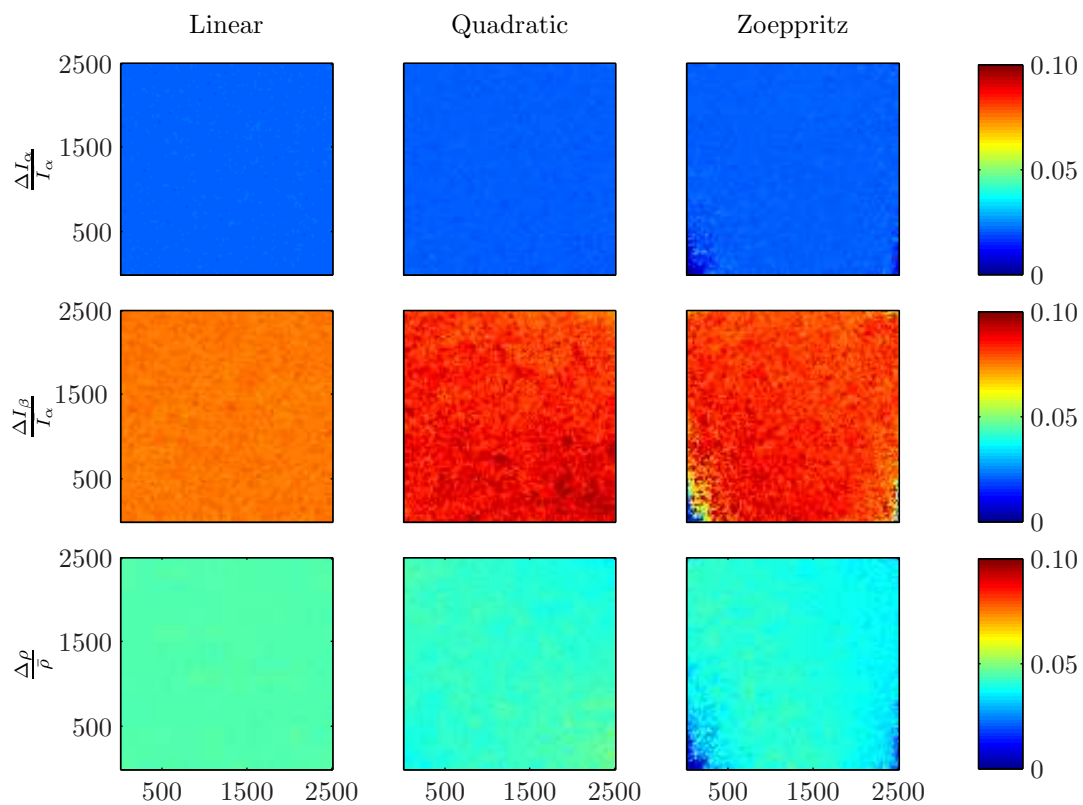
**Figure 7.** Standard deviation in the posterior distribution of  $\mathbf{m}$  from PP inversion without noise.



**Figure 8.** Posterior distribution of  $\sigma_m^2$  and  $\sigma_e^2$  from PP inversion without noise

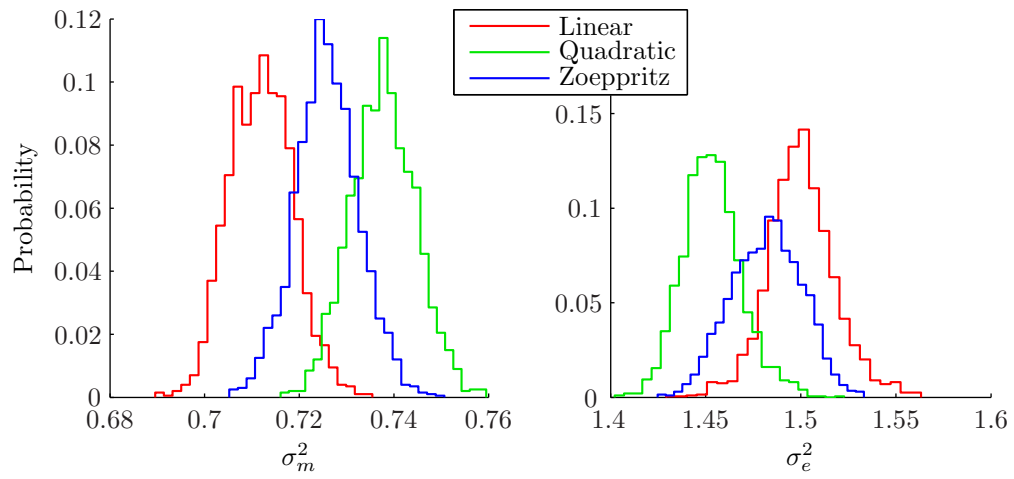


**Figure 9.** Bias in the posterior distribution of  $\mathbf{m}$  from PP inversion including multivariate Gaussian distributed noise.

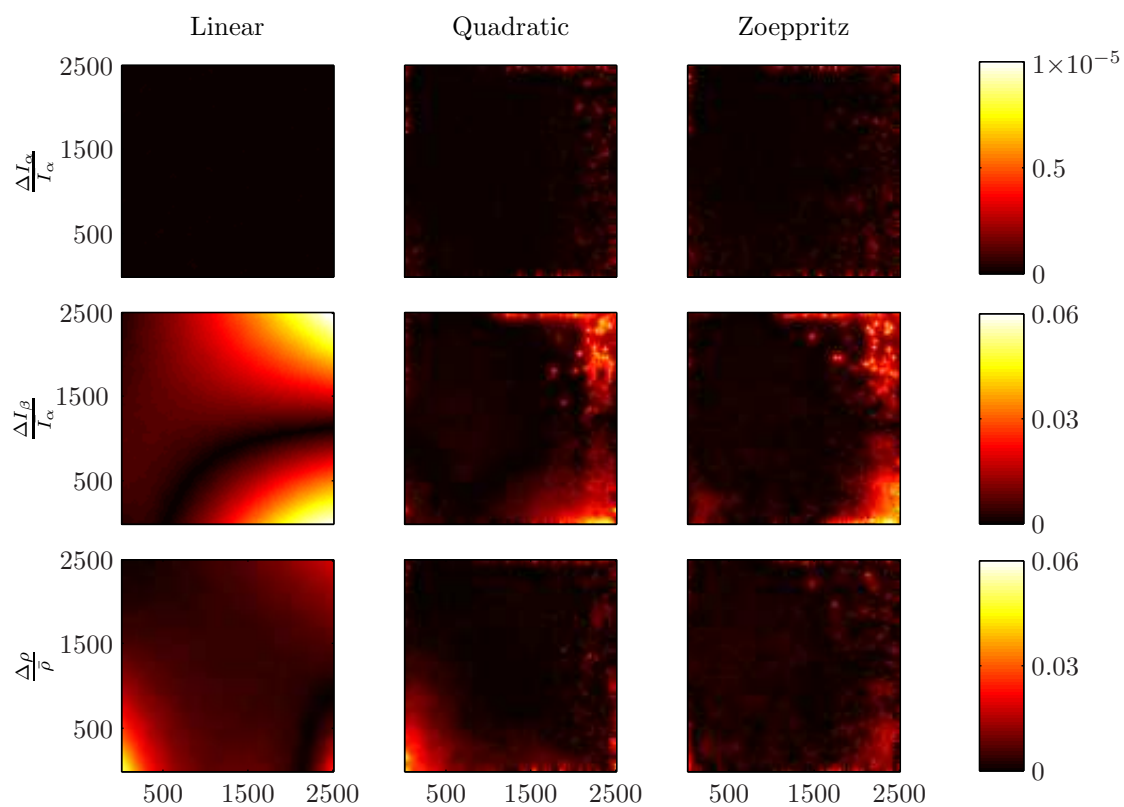


**Figure 10.** Standard deviation in the posterior distribution of  $\mathbf{m}$  from PP inversion including multivariate Gaussian distributed noise.

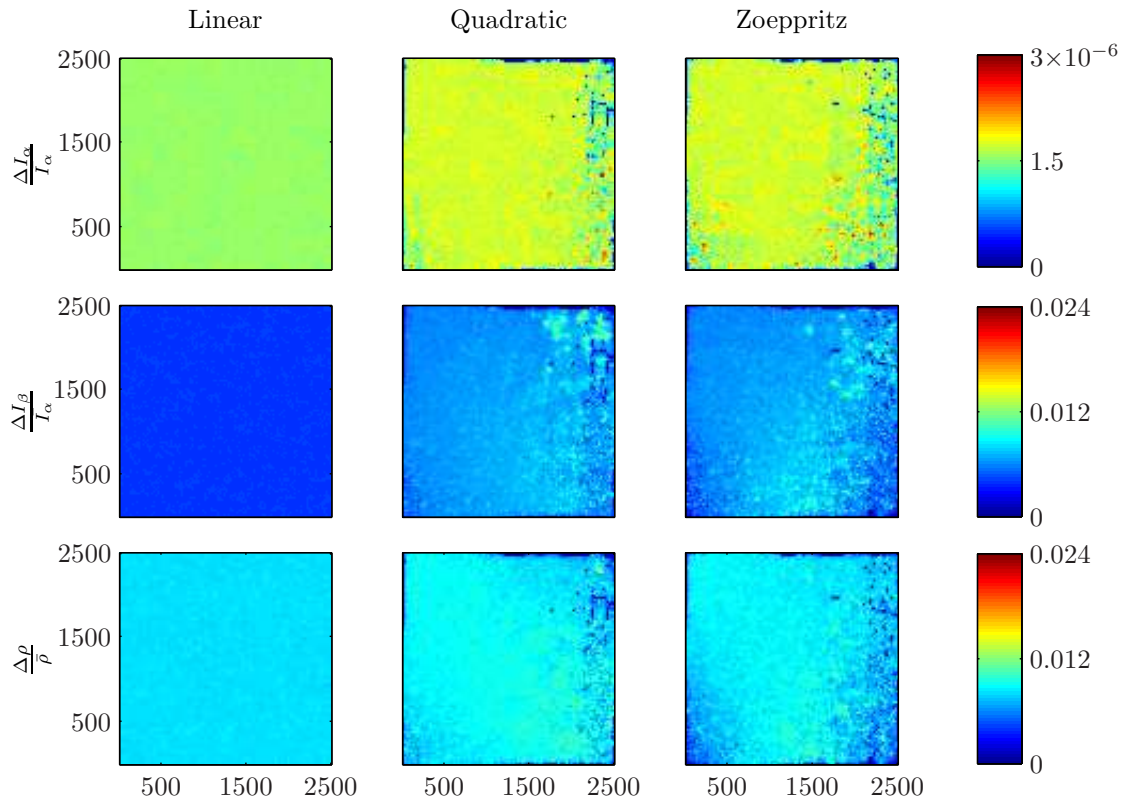




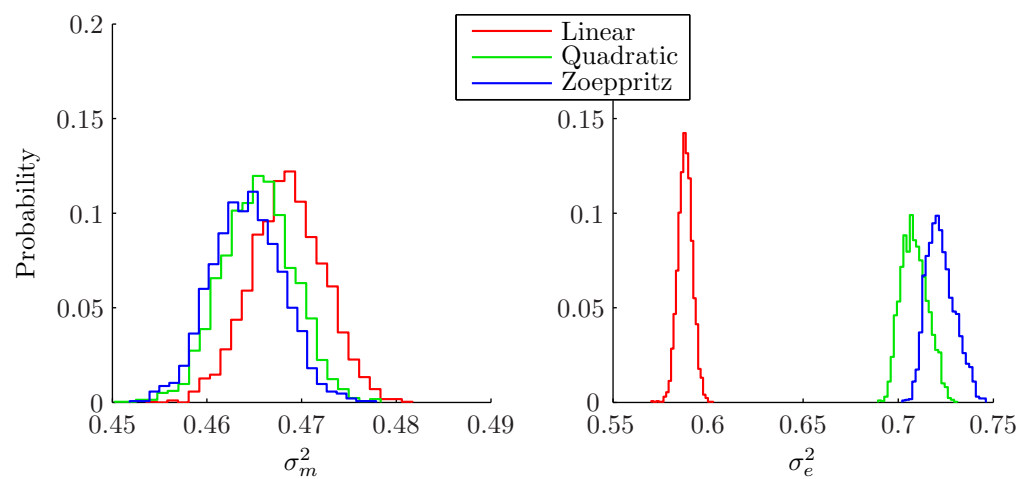
**Figure 11.** Posterior distribution of  $\sigma_m^2$  and  $\sigma_e^2$  from PP inversion including multivariate Gaussian distributed noise



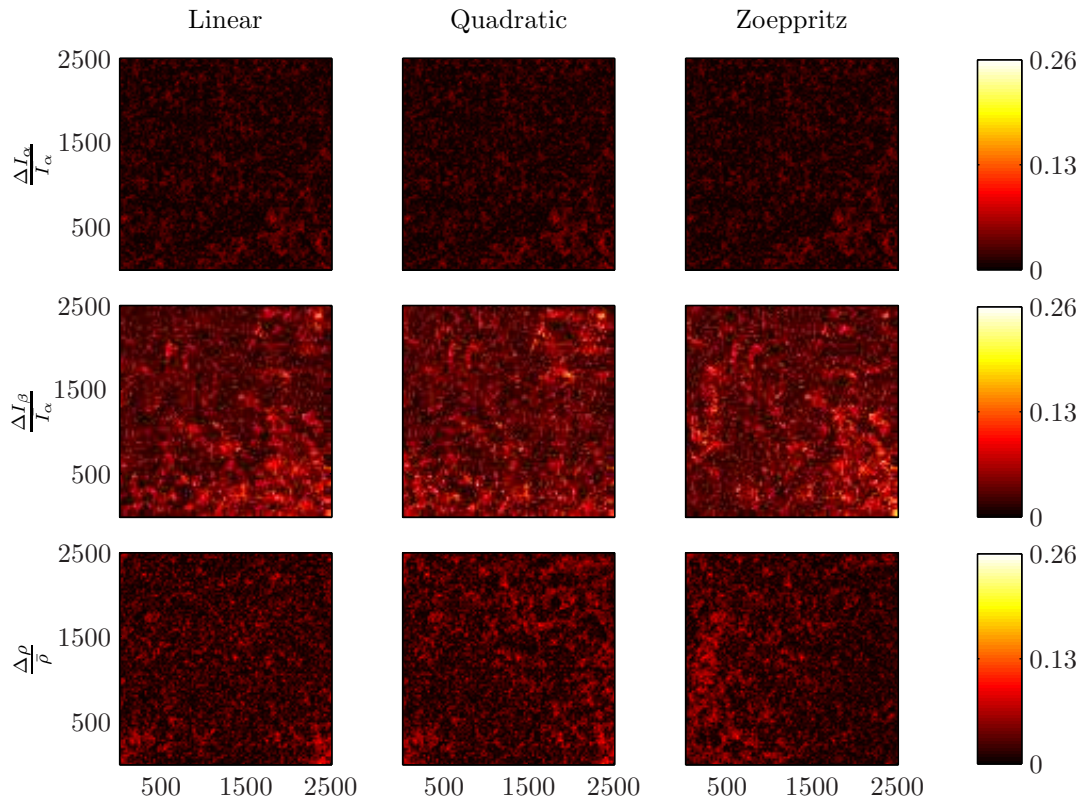
**Figure 12.** Bias in the posterior distribution of  $\mathbf{m}$  from joint PP and PS inversion without noise.



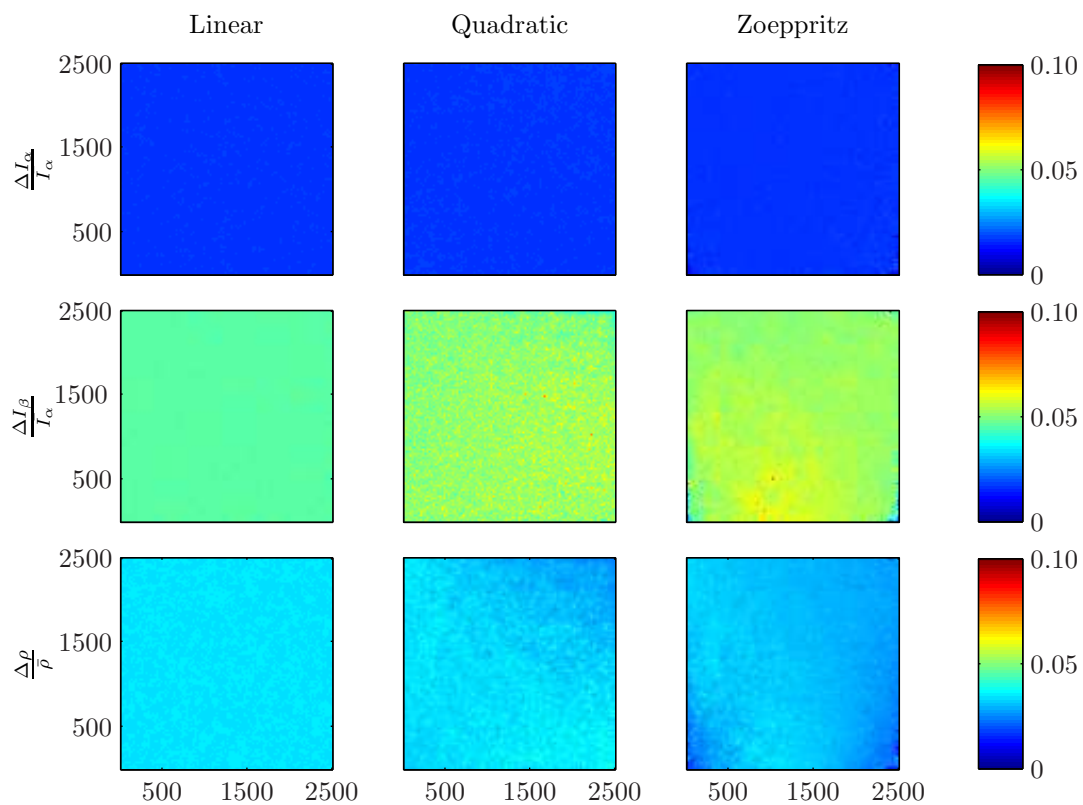
**Figure 13.** Standard deviation in the posterior distribution of  $\mathbf{m}$  from joint PP and PS inversion without noise.



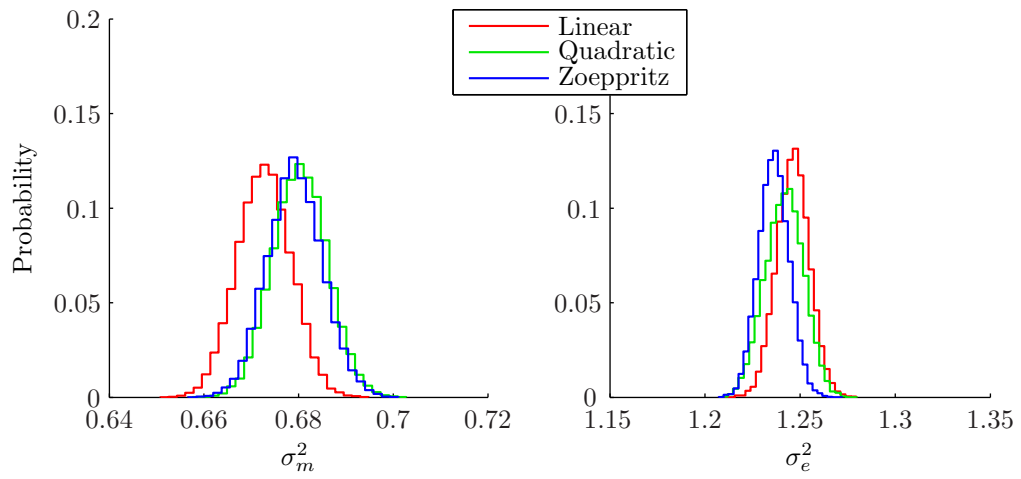
**Figure 14.** Posterior distribution of  $\sigma_m^2$  and  $\sigma_e^2$  from joint PP and PS inversion without noise



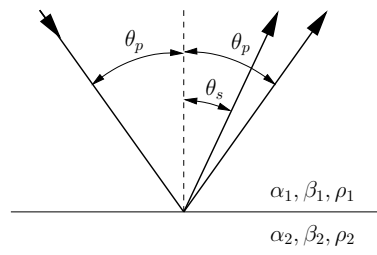
**Figure 15.** Bias in the posterior distribution of  $\mathbf{m}$  from joint PP and PS inversion including multivariate Gaussian distributed noise.



**Figure 16.** Standard deviation in the posterior distribution of  $\mathbf{m}$  from joint PP and PS inversion including multivariate Gaussian distributed noise.

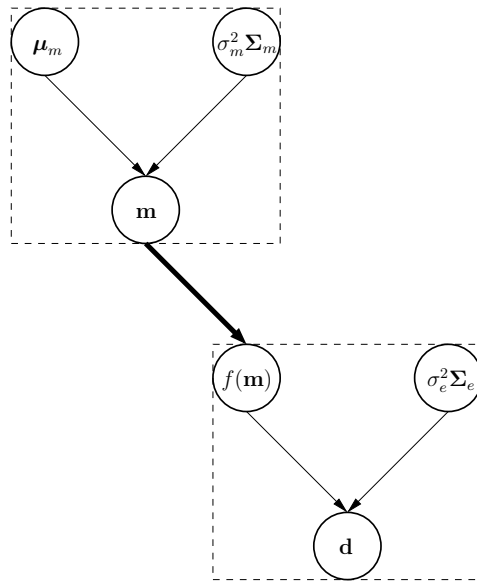


**Figure 17.** Posterior distribution of  $\sigma_m^2$  and  $\sigma_e^2$  from joint PP and PS inversion including multivariate Gaussian distributed noise

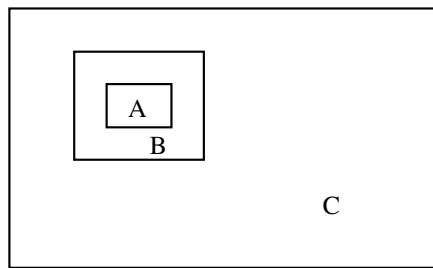


**Figure A1.** Relation between the downgoing P-wave  $\theta_p$ , reflected P-wave  $\theta_p$ , and reflected S-wave  $\theta_s$  and the definition of velocities and density in upper and lower medium.





**Figure B1.** Directed acyclic graph. The upper square is the prior model, the lower is the likelihood. The relation between  $\mathbf{m}$  and  $\mathbf{d}$  is the deterministic forward model.



**Figure C1.** Block A is the block to be updated, B is a boundary zone of limited thickness and C is the rest.