

# 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@ntnu.no*

## SUMMARY

Inversion of seismic reflection coefficients is 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. Quadratic approximations to the Zoeppritz equations are used as the forward model. Compared with the linear approximations the bias is reduced and the uncertainty estimate is 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 acceptance probability in the algorithm. Joint inversion using information from both reflected PP-waves and converted PS-waves yields smaller bias and reduced uncertainty 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 (Hilterman, 2001). 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 involve 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 lithology and fluid discrimination (Margrave et al., 2001; Sollid & Ursin, 2003).

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 squares approach to solve the inversion. Stewart (1990) and Veire & Landrø (2006) extended this to joint PP and PS inversion in a least squares 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 squares inversion and how to regularise it see Lines & Treitel (1984), and for a more comprehensive treatment of regularisation see Tenorio (2001).

Using a simplified model and incorporating all available data might not be enough to overcome the nonuniqueness problem and to 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 (Tarantola, 1987). It is very important to note that a Bayesian approach does not remove the nonuniqueness, it only helps identifying it.

Buland & Omre (2003a; 2003b) formulated the PP inversion, using a linear approximation, 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 about Bayesian modelling from a geophysics standpoint see Sen & Stoffa (1996), and from a statistical standpoint, see Robert & Casella (1999) and Liu (2001). Dubrule (2003) gives a general overview of geostatistical methods applied to seismic data.

Our approach to the inversion of reflection coefficients is to use isotropic, quadratic approximations 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 is 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 enables 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 (Mosegaard & Tarantola, 1995). 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 modelling and instead focus on the nonlinearity. The unknown model parameter vector,  $\mathbf{m}$ , contains the three elastic parameters in the Zoeppritz approximations. They are assumed to be multivariate Gaussian distributed with a (nonzero) prior mean and a prior covariance matrix which contains an unknown scaling factor,  $\sigma_m^2$ . The noise term in the data (PP or PP and PS reflection amplitudes) is also assumed multivariate Gaussian distributed, here with a zero prior mean and a prior noise covariance matrix which also contains an unknown scaling factor,  $\sigma_e^2$ . In the sampling procedure we invert for  $\mathbf{m}$ ,  $\sigma_m^2$ , and  $\sigma_e^2$  to produce a posterior distribution of these, given the data. The two scalars,  $\sigma_m^2$  and  $\sigma_e^2$ , are data driven and their main purpose is to stabilise the inversion algorithm, but they can to some extent quantify the noise level in the reflection coefficients.

In the following sections we first give explicit expressions for the 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 parameterization 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 (VTI) media. Explicit expressions for the PP and PS reflection coefficients, simplified for

two isotropic media, read

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

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

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, see Fig. 1.  $I_\alpha = \rho\alpha$  is P-wave impedance,  $I_\beta = \rho\beta$  is S-wave impedance and  $\rho$  is density.  $\Delta I_\alpha$  is difference between P-wave impedance in the lower and upper media and  $\bar{I}_\alpha$  denotes its average, similar definitions for  $I_\beta$  and  $\rho$ .

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

The variable to invert for, the medium parameters, is denoted  $\mathbf{m}$  and defined over a two dimensional  $n_x \times n_y$  lattice

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

where  $D_m$  is the number of medium parameters in each grid cell. With a linear approximation we have  $D_m = 3$ , i.e.  $m_{ij} = \left\{ \frac{\Delta I_\alpha}{\bar{I}_\alpha}, \frac{\Delta I_\beta}{\bar{I}_\beta}, \frac{\Delta \rho}{\bar{\rho}} \right\}_{ij}$ , see (1) and (2). The measurements,  $\mathbf{d}$ , are

defined for the same lattice,

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

where  $D_d$  is the number of measured reflection amplitudes (PP and/or PS) in each cell. We require  $D_d > D_m$  for the inverse problem not to be underdetermined.

The forward model is the link from  $\mathbf{m}$  to  $\mathbf{d}$  and can be written

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

where  $\mathbf{f}_z$  is the Zoeppritz equations and  $\mathbf{e}_z$  is the noise. The term  $\mathbf{e}_z$  consists of both measurement and modelling errors, but if the assumptions in the derivation of the Zoeppritz equation are valid, the modelling error part can be neglected.

Our focus is on the quadratic approximations and we then express the forward model as

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

where  $\mathbf{f}$ , in the case of both PP and PS reflections, is (1) and (2). If the quadratic approximations are good we will have  $\mathbf{e} \approx \mathbf{e}_z$ . We will also consider the linear forward model

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

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

A common way of inverting for  $\mathbf{m}$  is to minimise  $\mathbf{e}$  in (6) via an iterative nonlinear least squares algorithm. However, this produces no uncertainty measures, and we therefore reformulate the problem in a Bayesian context. The parameters and the measured data in (6) are then considered as stochastic variables and assigned probability distributions. In the Bayesian formulation the solution of the inverse problem is via Bayes' rule

$$\pi(\mathbf{m}|\mathbf{d}) = \frac{\pi(\mathbf{d}|\mathbf{m})\pi(\mathbf{m})}{\pi(\mathbf{d})} \propto \pi(\mathbf{d}|\mathbf{m})\pi(\mathbf{m}), \quad (8)$$

where we use  $\pi$  as a generic symbol for any probability distribution. The posterior distribution  $\pi(\mathbf{m}|\mathbf{d})$  is the distribution of the model parameters  $\mathbf{m}$  given the observed data. This is the distribution we would like to assess. The first factor in the numerator,  $\pi(\mathbf{d}|\mathbf{m})$ , is

known as the likelihood function and corresponds to the forward model discussed above. The second factor in the numerator,  $\pi(\mathbf{m})$ , is known as the prior distribution and should specify available information about  $\mathbf{m}$  before the data  $\mathbf{d}$  is measured. This should include both theoretical knowledge, e.g. velocities can not be negative, and case dependent knowledge, e.g. information from the geological setting. Its main purpose is to restrict the model parameter space to reasonable values. The denominator in (8),  $\pi(\mathbf{d}) = \int \pi(\mathbf{d}|\mathbf{m}) \pi(\mathbf{m}) d\mathbf{m}$ , is the (prior) marginal distribution of  $\mathbf{d}$ . Once the data  $\mathbf{d}$  is observed it is just a normalising constant that is of no interest to us, therefore the proportionality in (8).

To use (8) we first have to assign the likelihood and prior distributions. We start by assuming the prior distributions for the model parameters and the noise to have multivariate normal distributions

$$\pi(\mathbf{m}) = \mathcal{N}(\mathbf{m}; \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \quad (9)$$

$$\pi(\mathbf{e}) = \mathcal{N}(\mathbf{e}; 0, \boldsymbol{\Sigma}_e), \quad (10)$$

where  $\boldsymbol{\mu}_m$  is the prior expected value of  $\mathbf{m}$ , and  $\boldsymbol{\Sigma}_m$  and  $\boldsymbol{\Sigma}_e$  are the prior and noise covariance matrices, respectively. For details about the multivariate Gaussian distribution see Appendix C. Combining (6) and (10) and assuming a deterministic forward model  $\mathbf{f}$ , the likelihood function becomes

$$\pi(\mathbf{d}|\mathbf{m}) = \mathcal{N}(\mathbf{d}; \mathbf{f}(\mathbf{m}), \boldsymbol{\Sigma}_e). \quad (11)$$

The covariance matrices  $\boldsymbol{\Sigma}_m$  and  $\boldsymbol{\Sigma}_e$  decide the influence of the prior and the data, respectively, on the resulting posterior distribution. If the variance of  $\mathbf{e}$  is small compared to the prior variance, the posterior will be mainly determined by the data  $\mathbf{d}$ , whereas if the prior variance is very small compared to the variance of  $\mathbf{e}$ , the posterior distribution will be concentrated around the prior mean  $\boldsymbol{\mu}_m$ . The specification of the two covariance matrices is therefore important. We partly let this specification be decided by the data by setting

$$\boldsymbol{\Sigma}_m = \sigma_m^2 \mathbf{S}_m \quad (12)$$

$$\boldsymbol{\Sigma}_e = \sigma_e^2 \mathbf{S}_e, \quad (13)$$

where  $\mathbf{S}_m$  and  $\mathbf{S}_e$  are a priori given matrices, and the estimation of the scalars  $\sigma_m^2$  and  $\sigma_e^2$  is part of the inversion procedure. In this formulation we have to specify only the structure of the two covariance matrices. The new prior and likelihood distributions now read

$$\pi(\mathbf{m}|\sigma_m^2) = \mathcal{N}(\mathbf{m}; \boldsymbol{\mu}_m, \sigma_m^2 \mathbf{S}_m) \quad (14)$$

$$\pi(\mathbf{d}|\mathbf{m}, \sigma_e^2) = \mathcal{N}(\mathbf{d}; \mathbf{f}(\mathbf{m}), \sigma_e^2 \mathbf{S}_e). \quad (15)$$

In addition we now need to specify prior distributions for the scalars  $\sigma_e^2$  and  $\sigma_m^2$ . We adopt inverse gamma distributions,

$$\pi(\sigma_e^2) = \mathcal{IG}(\sigma_e^2; \alpha_e, \beta_e) \quad (16)$$

$$\pi(\sigma_m^2) = \mathcal{IG}(\sigma_m^2; \alpha_m, \beta_m), \quad (17)$$

where  $\alpha_e$ ,  $\beta_e$ ,  $\alpha_m$  and  $\beta_m$  are scalar constants. The inverse gamma distribution is a flexible model defined for positive values and can thereby be adapted to different prior knowledge. Moreover, it makes the mathematical treatment of the resulting posterior distributions easier. For details about the inverse gamma distribution see Appendix C.

Including  $\sigma_e^2$  and  $\sigma_m^2$  as unknown variables together with  $\mathbf{m}$ , the posterior distribution becomes

$$\pi(\mathbf{m}, \sigma_m^2, \sigma_e^2 | \mathbf{d}) \propto \pi(\mathbf{d} | \mathbf{m}, \sigma_m^2, \sigma_e^2) \pi(\mathbf{m}, \sigma_m^2, \sigma_e^2) = \pi(\mathbf{d} | \mathbf{m}, \sigma_e^2) \pi(\mathbf{m} | \sigma_m^2) \pi(\sigma_m^2) \pi(\sigma_e^2), \quad (18)$$

where we have assumed  $\sigma_e^2$  and  $\sigma_m^2$  to be a priori independent,  $\mathbf{m}$  to be conditionally independent of  $\sigma_e^2$  given  $\sigma_m^2$ , and  $\mathbf{d}$  to be conditionally independent of  $\sigma_m^2$  given  $\mathbf{m}$  and  $\sigma_e^2$  (all implicitly assumed in (14)-(17)). Fig. 2 summarises our model and displays a graphical representation of the input to and output from the inversion algorithm.

### 3 INVERSION ALGORITHM

Our knowledge about  $\mathbf{m}$  after having observed the data  $\mathbf{d}$  is described by the posterior distribution

$$\pi(\mathbf{m} | \mathbf{d}) = \int \int \pi(\mathbf{m}, \sigma_e^2, \sigma_m^2 | \mathbf{d}) d\sigma_e^2 d\sigma_m^2. \quad (19)$$

The two commonly used estimators for estimating the unknown  $\mathbf{m}$  is the posterior mode (or the maximum a posteriori (MAP) estimate)

$$\arg \max_{\mathbf{m}} \pi(\mathbf{m}|\mathbf{d}), \quad (20)$$

and the posterior mean

$$\boldsymbol{\mu}_{\mathbf{m}|\mathbf{d}} = \mathbb{E}[\mathbf{m}|\mathbf{d}] = \int \mathbf{m} \pi(\mathbf{m}|\mathbf{d}) \, d\mathbf{m}. \quad (21)$$

For Gaussian posterior distributions the two estimators coincide. Our nonlinear (quadratic) forward function  $\mathbf{f}(\mathbf{m})$  makes  $\pi(\mathbf{m}|\mathbf{d})$  non-Gaussian and the two estimators will give different results. For  $\boldsymbol{\mu}_{\mathbf{m}|\mathbf{d}}$  the associated uncertainty is easily quantified by the associated standard deviations, and for this reason we focus on  $\boldsymbol{\mu}_{\mathbf{m}|\mathbf{d}}$  in the following.

The posterior distributions  $\pi(\mathbf{m}, \sigma_e^2, \sigma_m^2 | \mathbf{d})$  and  $\pi(\mathbf{m} | \mathbf{d})$  are not analytically tractable due to the nonlinear  $\mathbf{f}(\mathbf{m})$ . The posterior mode estimator (20) must be found by numerical optimisation, whereas  $\boldsymbol{\mu}_{\mathbf{m}|\mathbf{d}}$  can be found via stochastic simulation. For the latter, the idea is to generate a (large) number of realisations,  $\mathbf{m}^{(1)}, \dots, \mathbf{m}^{(K)}$  from  $\pi(\mathbf{m}|\mathbf{d})$  and estimate the posterior mean  $\boldsymbol{\mu}_{\mathbf{m}|\mathbf{d}}$  by

$$\hat{\boldsymbol{\mu}}_{\mathbf{m}|\mathbf{d}} = \frac{1}{K} \sum_{k=1}^K \mathbf{m}^{(k)}. \quad (22)$$

The associated standard deviation is easily estimated from the same realisations. Direct sampling from  $\pi(\mathbf{m}|\mathbf{d})$  is not possible, again due to the nonlinear  $\mathbf{f}(\mathbf{m})$ . Instead we adopt the Metropolis–Hastings algorithm (Robert & Casella, 1999; Liu, 2001). In the following we give a brief introduction to this algorithm and how it is used in our problem. More detailed general discussions can be found in the references given, and for our situation in Appendices A and B.

We use the Metropolis–Hastings algorithm to generate samples from  $\pi(\mathbf{m}, \sigma_e^2, \sigma_m^2 | \mathbf{d})$ . The generated  $\mathbf{m}$ 's are then automatically also realisations from  $\pi(\mathbf{m}|\mathbf{d})$ . The Metropolis–Hastings algorithm is iterative. It first assigns some initial values to the state vector  $(\mathbf{m}, \sigma_e^2, \sigma_m^2)$  and then starts iterating. Each iteration consists of two parts, first potential new values for the state vector (or parts of it) are proposed according to a proposal distribution. Second,

the potential new values are accepted with a specific acceptance probability that depends on the current values of the state vector, the proposed new values, the posterior distribution, and the proposal distribution used. If the proposed values are not accepted, the old ones are retained. The main idea of the algorithm is that the acceptance probability is constructed so that in the limit, when the number of iterations goes to infinity, the probability distribution of the state vector equals the target distribution  $\pi(\mathbf{m}, \sigma_e^2, \sigma_m^2 | \mathbf{d})$ . Thus, after a sufficiently large number of iterations the simulated values  $(\mathbf{m}, \sigma_e^2, \sigma_m^2)$  are (essentially) distributed according to  $\pi(\mathbf{m}, \sigma_e^2, \sigma_m^2 | \mathbf{d})$  and the procedure thereby produces the necessary realisations to estimate  $\boldsymbol{\mu}_{\mathbf{m}|\mathbf{d}}$  and corresponding standard deviations.

In our implementation of the Metropolis–Hastings algorithm we alternate between three proposal distributions, each proposing a change in one of  $\sigma_e^2$ ,  $\sigma_m^2$ , and  $\mathbf{m}$  and keeping the other two unchanged. The first of the three proposes a potential new value for  $\sigma_e^2$  by generating it from  $\pi(\sigma_e^2 | \mathbf{d}, \mathbf{m}, \sigma_m^2)$ , where the current values for  $\mathbf{m}$  and  $\sigma_m^2$  are used. This is known as the full conditional distribution for  $\sigma_e^2$  and is an analytically tractable distribution in our model, see Appendix A for details. To use the full conditional distribution as a proposal distribution is also known as a Gibbs step and the acceptance probability for such a proposal always equals one. In the second proposal distribution the full conditional distribution for  $\sigma_m^2$ ,  $\pi(\sigma_m^2 | \mathbf{d}, \mathbf{m}, \sigma_e^2)$ , is used to propose a potential new value for  $\sigma_m^2$ . Again the full conditional is an analytically tractable distribution, and the acceptance probability is one. The third proposal distribution generates potential new values for (a block in)  $\mathbf{m}$ . The full conditional for  $\mathbf{m}$  is not analytically tractable in our model, because of the nonlinear  $\mathbf{f}(\mathbf{m})$ , and can therefore not be used as a proposal distribution. Instead we consider a linearized model where a linear  $\mathbf{F}\mathbf{m}$  is substituted for  $\mathbf{f}(\mathbf{m})$ , see the discussion in Section 2. For this linearized model the full conditional distribution for  $\mathbf{m}$  is Gaussian and can be used as a proposal distribution. It should be noted that the approximation done when using a linearized  $\mathbf{f}(\mathbf{m})$  is corrected for in the acceptance probability of the Metropolis–Hastings algorithm, so the generated realisations are really distributed according to the desired  $\pi(\mathbf{m}, \sigma_e^2, \sigma_m^2 | \mathbf{d})$ . However, the approximation done in the proposal distribution has a price, the acceptance probability will

often be less than one, and as a consequence, more iterations must be run before convergence. The larger the approximation, the lower the acceptance probability typically becomes. To avoid a very low acceptance probability we therefore only propose new values for a block of cells in  $\mathbf{m}$ , see Appendix B for details.

#### 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: only PP data and joint PP and PS data, both of them with and without noise added. For all these four cases we will compare inversions with three different forward models: linear, quadratic, and exact Zoeppritz.

The truth we have chosen is, as discussed earlier, parameterized 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. 3. We are using a model which is smoothly spatially varying because this is a likely scenario at an interface as long as there are no faults or changes in fluid content. The contrasts are very strong, and the purpose is to test the nonlinear inversion algorithm. With this choice we also try to span the model space to see how the inversion performs with combinations of parameters with different magnitudes. Of course, in order to fully span the model space, we would have to use a three dimensional space. 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 we use the linear forward model 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 again leads to poor convergence and run time problems. A reason for introducing the two scalars  $\sigma_e^2$  and  $\sigma_m^2$  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 in the inversion. It quantifies how the model can reproduce  $\mathbf{d}$  from the posterior  $\mathbf{m}$ , but this does not necessarily quantify the bias.

The expected value of the prior of  $\mathbf{m}$  is in all cases half of the true  $\mathbf{m}$ , and the reason is to see if the inversion relies too heavily on prior information. If this is the case, the end result will appear with a large bias towards the prior information.

Our first set of measurements is PP reflections without any noise. Fig. 4 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 omitting 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 one has large bias.

After inverting PP reflections we will turn to joint PP and PS inversion. We will use the four coefficients in Fig. 4 and include three PS reflections with incoming P-waves between  $20^\circ$  and  $55^\circ$  to see what additional information the PS reflections will bring. Fig. 5 displays the exact Zoeppritz reflection coefficients together with the bias in the approximations, similar to the PP case. Here it is even clearer how superior the quadratic approximations are.

Next, we will test PP inversion when noise is included. In the prior of  $\mathbf{m}$  we add spatially correlated normally distributed noise to break the smoothness, and to the measurements  $\mathbf{d}$  we add normally distributed noise, correlated both spatially and between angles. In the left column of Fig. 6 we see the PP reflection data 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.

The last synthetic example is joint inversion of reflection coefficients including noise. The prior used is the same as in the previous case. For the measurements the situation is almost equal. The variance is 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. 6 shows the measurements generated for joint inversion of PP and PS reflections including noise.

## 5 NUMERICAL RESULTS

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 therefore simple to visualise, but this is not the case for  $\mathbf{m}$  since it is multivariate. We therefore calculate the posterior mean (21) and subtract the truth to produce the bias. Fig. 7 shows the absolute value of this bias in the case without noise. Second, as mentioned in Section 3, we can generate uncertainty estimates from the realisations, and in Fig. 8 the standard deviation of the distribution is plotted. Finally, the distributions of  $\sigma_e^2$  and  $\sigma_m^2$  are displayed in Fig. 9.

The second example is joint inversion without noise added. As for the previous PP case 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 10 and 11 show the bias and the standard deviation and Fig. 12 the two corresponding scalar distributions.

Case number three that we tested was PP data including noise. The absolute value of the bias of the mean is displayed in Fig. 13, the standard deviation is in Fig. 14, and the scalar distributions are in Fig. 15.

Figs 16, 17, and 18 show the same plots for the last case, joint inversion of PP and PS data with noise added.

## 6 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  $\sigma_m^2$ . This section will end with some remarks on how to proceed in the case of real seismic data.

The absolute value of the bias in PP inversion and joint inversion, both without noise,

found in Figs 7 and 10, have interesting common features and differences. Because of the parameterization 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. 5 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 8 and 11. 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 is much smaller compared with the two other parameters, as for the bias. For the two linear models the standard 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. 8 the linear inversion shows 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 uncertainties in the nonlinear inversions yield 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 13 and 16. 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 14 and 17 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.

To conclude the discussion of the numerical results we will briefly comment on the two scalar parameters,  $\sigma_e^2$  and  $\sigma_m^2$ . As mentioned earlier, the main reason for including these is to stabilise 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.

An interesting question that arises is how this algorithm will perform on real seismic data. Certainly there are several problems to overcome. First we have the strength of the contrasts, if they are very weak the linear approximations will work fine. Second we have the quality of the angle gathers. In order to exploit the high accuracy of the quadratic forward model, the AVA gathers need a high signal-to-noise ratio. A way to achieve this is to carefully choose a processing workflow that keeps as much as possible of the amplitude information, and avoid steps like automatic gain control (AGC). In the case of PS data we also have to handle converted wave processing and estimate a good S-wave velocity model. However, even with this careful processing we still need to scale globally the amplitudes using well log or geological information. In the inversion procedure we also have to determine the covariance matrices  $\mathbf{S}_m$  and  $\mathbf{S}_e$  together with the prior mean  $\boldsymbol{\mu}_m$ . However, all these issues are related to the example of inversion of seismic reflection amplitudes. The nonlinear inversion algorithm can most certainly be applied to other nonlinear problems.

## 7 CONCLUSION

In our approach to the inversion of seismic reflection coefficients we have formulated 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 modelling. We have also used new

quadratic approximations to the Zoeppritz equations and compared them with both the linearized 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.
- Dubrule, O., 2003. *Geostatistics for Seismic Data Integration in Earth Models*, Distinguished Instructor Series, No. 6, Society of Exploration Geophysicists.
- Hampson, D., 1991. AVO inversion, theory and practice, *The Leading Edge*, **10**, 39–42.
- Hilterman, F. J., 2001. *Seismic Amplitude Interpretation*, Distinguished Instructor Series, No. 4, Society of Exploration Geophysicists.

- 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*, **20**, 1048–1052.
- Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *Journal of Geophysical Research*, **100**, 12,431–12,447.
- 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.
- Sollid, A. & Ursin, B., 2003. Scattering-angle migration of ocean-bottom seismic data in weakly anisotropic media, *Geophysics*, **68**, 641–655.
- 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 viscoelastic 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.
- Tarantola, A., 1987. *Inverse Problem Theory*, Elsevier.
- 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: SAMPLING

Due to the nonlinear forward model,  $\mathbf{f}(\mathbf{m})$ , analytical calculations of the posterior distribution (18) is not possible. However, realisations from the posterior can be generated by the Metropolis–Hastings algorithm (Robert & Casella (1999) and Liu (2001)). This algorithm is iterative, and the probability distribution of the variables that are simulated converges

to the posterior distribution of interest as the number of iterations goes to infinity. Each iteration consists of two steps. First, potential new values for the variables to be simulated are generated from a *proposal distribution*. Second, the potential new values are accepted with an *acceptance probability*. The proposal distribution can be chosen quite freely. It does not influence the limiting probability distribution, but is crucial for the convergence rate of the algorithm. For a general discussion of the algorithm see the references given above. In the following we discuss our choice of proposal distributions and resulting acceptance probabilities for the posterior distribution (18).

We use three proposal distributions, one proposal distribution for each of  $\sigma_e^2$  and  $\sigma_m^2$  and one for  $\mathbf{m}$ . With our choice of likelihood and priors, (15), (16), and (17), both  $\pi(\sigma_e^2|\mathbf{d}, \mathbf{m}, \sigma_m^2)$  and  $\pi(\sigma_m^2|\mathbf{d}, \mathbf{m}, \sigma_e^2)$  become inverse gamma distributions, see Appendix C for a detailed proof,

$$\begin{aligned} \pi(\sigma_e^2|\mathbf{d}, \mathbf{m}, \sigma_m^2) &= \mathcal{IG}\left(\sigma_e^2 \left| \alpha_e + \frac{n_e}{2}, \beta_e + s_e^2 \frac{n_e}{2} \right.\right) \\ \pi(\sigma_m^2|\mathbf{d}, \mathbf{m}, \sigma_e^2) &= \mathcal{IG}\left(\sigma_m^2 \left| \alpha_m + \frac{n_m}{2}, \beta_m + s_m^2 \frac{n_m}{2} \right.\right), \end{aligned} \tag{A1}$$

where

$$\begin{aligned} s_e^2 &= \frac{1}{n_e}(\mathbf{d} - \mathbf{f}(\mathbf{m}))^T \mathbf{S}_e^{-1}(\mathbf{d} - \mathbf{f}(\mathbf{m})) \\ s_m^2 &= \frac{1}{n_m}(\mathbf{m} - \boldsymbol{\mu}_m)^T \mathbf{S}_m^{-1}(\mathbf{m} - \boldsymbol{\mu}_m), \end{aligned} \tag{A2}$$

$n_e = n_x n_y D_d$  and  $n_m = n_x n_y D_m$ . Inverse gamma distributions are analytical tractable and we therefore use  $\pi(\sigma_e^2|\mathbf{d}, \mathbf{m}, \sigma_m^2)$  and  $\pi(\sigma_m^2|\mathbf{d}, \mathbf{m}, \sigma_e^2)$  as proposal distributions for  $\sigma_e^2$  and  $\sigma_m^2$ , respectively. Proposal distributions like this, where the potential new value for a variable is generated from the full conditional for this variable given everything else, i.e. data and all other variables, is known as a Gibbs step (Robert & Casella (1999) and Liu (2001)), and the acceptance probability is then always equal to one. Thus, the acceptance step can be ignored when generating new values for  $\sigma_e^2$  and  $\sigma_m^2$ .

As  $\pi(\mathbf{m}|\mathbf{d}, \sigma_e^2, \sigma_m^2)$  is not analytical tractable, a Gibbs step can not be used for  $\mathbf{m}$ . It is the nonlinear  $\mathbf{f}(\mathbf{m})$  that prohibit analytical calculations for  $\pi(\mathbf{m}|\mathbf{d}, \sigma_e^2, \sigma_m^2)$ . A natural alternative is therefore instead to consider the linear likelihood in (7) and the corresponding

resulting full conditional

$$\pi^{lin}(\mathbf{m}|\mathbf{d}, \sigma_e^2, \sigma_m^2) \propto \pi^{lin}(\mathbf{d}|\mathbf{m}, \sigma_e^2) \pi(\mathbf{m}|\sigma_m^2), \quad (\text{A3})$$

where  $\pi^{lin}(\mathbf{d}|\mathbf{m}, \sigma_e^2)$  is the likelihood function corresponding to (7). The  $\pi^{lin}(\mathbf{m}|\mathbf{d}, \sigma_e^2, \sigma_m^2)$  is Gaussian and thereby easy to sample from. In this respect  $\pi^{lin}(\mathbf{m}|\mathbf{d}, \sigma_e^2, \sigma_m^2)$  could have been used as a proposal distribution for  $\mathbf{m}$ . However, the acceptance probabilities then become very low. Loosely speaking, the approximation induced by proposing new values from  $\pi^{lin}(\mathbf{m}|\mathbf{d}, \sigma_e^2, \sigma_m^2)$  instead of from the correct full conditional  $\pi(\mathbf{m}|\mathbf{d}, \sigma_e^2, \sigma_m^2)$  is corrected for by the acceptance probability. When simulating  $\mathbf{m}$  on a large lattice the difference between  $\pi^{lin}(\mathbf{m}|\mathbf{d}, \sigma_e^2, \sigma_m^2)$  and  $\pi(\mathbf{m}|\mathbf{d}, \sigma_e^2, \sigma_m^2)$  becomes substantial. A consequence is very small acceptance probabilities, which in turn gives very slow convergence of the algorithm. To overcome this problem we instead propose new values for only a block  $A$  of  $\mathbf{m}$  at a time, see Fig. 19. More precisely, our proposal distribution for block  $A$  in  $\mathbf{m}$  is

$$\pi^{lin}(\mathbf{m}_A|\mathbf{m}_B, \mathbf{d}_A, \mathbf{d}_B, \sigma_m^2, \sigma_e^2), \quad (\text{A4})$$

where  $\mathbf{m}_A$  and  $\mathbf{m}_B$  denote the parts of  $\mathbf{m}$  inside the sets  $A$  and  $B$ , respectively, and  $\mathbf{d}_A$  and  $\mathbf{d}_B$  are the corresponding parts of the data vector  $\mathbf{d}$ . The acceptance probability becomes

$$p(\tilde{\mathbf{m}}|\mathbf{m}) = \min \left\{ 1, \frac{\pi(\tilde{\mathbf{m}}|\mathbf{d}, \sigma_e^2, \sigma_m^2)}{\pi(\mathbf{m}|\mathbf{d}, \sigma_e^2, \sigma_m^2)} \cdot \frac{\pi^{lin}(\mathbf{m}_A|\mathbf{m}_B, \mathbf{d}_A, \mathbf{d}_B, \sigma_e^2, \sigma_m^2)}{\pi^{lin}(\tilde{\mathbf{m}}_A|\mathbf{m}_B, \mathbf{d}_A, \mathbf{d}_B, \sigma_e^2, \sigma_m^2)} \right\}, \quad (\text{A5})$$

where  $\mathbf{m} = (\mathbf{m}_A, \mathbf{m}_B, \mathbf{m}_C)$ ,  $\tilde{\mathbf{m}} = (\tilde{\mathbf{m}}_A, \mathbf{m}_B, \mathbf{m}_C)$ , and  $\mathbf{m}_A$  and  $\tilde{\mathbf{m}}_A$  are the old and the proposed new values for the variable  $\mathbf{m}_A$ , respectively. Details on  $\pi^{lin}(\mathbf{m}_A|\mathbf{m}_B, \mathbf{d}_A, \mathbf{d}_B, \sigma_m^2, \sigma_e^2)$  are discussed in Appendix B. Loosely speaking, by proposing new values for  $\mathbf{m}$  only in the block  $A$ , the resulting error by using the linearized forward model in the proposal distribution is reduced, and the acceptance probability is correspondingly increased. In turn, this gives better convergence properties for the simulation algorithm. It should be noted that in the proposal distribution (A4) we are conditioning on values of  $\mathbf{m}$  in the boundary zone  $B$ ,  $\mathbf{m}_B$ . By this we impose the necessary continuity between the values in  $B$  and the new values in  $A$ . We do not condition on  $\mathbf{m}_C$  in the proposal distribution as this would dramatically

increase the computer resources necessary to generate the proposals, but with only a minor influence on the proposed values.

To summarise, our simulation algorithm becomes

- (i) Initiate values for  $\sigma_e^2$ ,  $\sigma_m^2$  and  $\mathbf{m}$ .
- (ii) Iterate for a large number of iterations.
  - (a) Generate new values for  $\sigma_e^2$  and  $\sigma_m^2$  by sampling according to the distributions in (A1).
  - (b) For a series of (overlapping) blocks  $A$ , covering the whole lattice used,
    - 1. Propose new values for  $\mathbf{m}_A$ ,  $\tilde{\mathbf{m}}_A$ , by sampling from (A4).
    - 2. Compute the acceptance probability  $p(\tilde{\mathbf{m}}|\mathbf{m})$  given by (A5).
    - 3. Generate  $u \in [0, 1]$  from a uniform distribution on the unit interval.
    - 4. If  $u \leq p(\tilde{\mathbf{m}}|\mathbf{m})$  accept the proposed new values, i.e. set  $\mathbf{m}_A = \tilde{\mathbf{m}}_A$ , otherwise retain the old values  $\mathbf{m}_A$ .

## APPENDIX B: LINEAR PROPOSAL DISTRIBUTION

To sample  $\tilde{\mathbf{m}}_A$  from the distribution (A4) we consider the linear forward model (7). For the two blocks  $A$  and  $B$  it reads

$$\mathbf{d}_A = \mathbf{F}_A \mathbf{m}_A + \mathbf{e}_A, \tag{B1}$$

$$\mathbf{d}_B = \mathbf{F}_B \mathbf{m}_B + \mathbf{e}_B.$$

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

$$\begin{bmatrix} \mathbf{m}_A \\ \mathbf{m}_B \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{bmatrix}, \begin{bmatrix} \sigma_m^2 \mathbf{S}_{m,AA} & \sigma_m^2 \mathbf{S}_{m,AB} \\ \sigma_m^2 \mathbf{S}_{m,BA} & \sigma_m^2 \mathbf{S}_{m,BB} \end{bmatrix} \right) \tag{B2}$$

and

$$\begin{bmatrix} \mathbf{e}_A \\ \mathbf{e}_B \end{bmatrix} \sim \mathcal{N} \left( 0, \begin{bmatrix} \sigma_e^2 \mathbf{S}_{e,AA} & \sigma_e^2 \mathbf{S}_{e,AB} \\ \sigma_e^2 \mathbf{S}_{e,BA} & \sigma_e^2 \mathbf{S}_{e,BB} \end{bmatrix} \right). \tag{B3}$$

From these the mean vectors and covariance matrices between the different  $\mathbf{m}_A$ ,  $\mathbf{m}_B$ ,  $\mathbf{d}_A$ , and  $\mathbf{d}_B$  can easily be found. For instance

$$\begin{aligned} \mathbb{E}(\mathbf{d}_A) &= \mathbb{E}(\mathbf{F}_A \mathbf{m}_A + \mathbf{e}_A) \\ &= \mathbf{F}_A \boldsymbol{\mu}_A, \end{aligned} \tag{B4}$$

$$\begin{aligned} \text{Cov}(\mathbf{m}_A, \mathbf{d}_B) &= \text{Cov}(\mathbf{m}_A, \mathbf{F}_B \mathbf{m}_B + \mathbf{e}_B) \\ &= \sigma_m^2 \mathbf{S}_{m,AB} \mathbf{F}_B^T, \end{aligned} \tag{B5}$$

$$\begin{aligned} \text{Cov}(\mathbf{d}_A, \mathbf{d}_B) &= \text{Cov}(\mathbf{F}_A \mathbf{m}_A + \mathbf{e}_A, \mathbf{F}_B \mathbf{m}_B + \mathbf{e}_B) \\ &= \text{Cov}(\mathbf{F}_A \mathbf{m}_A, \mathbf{F}_B \mathbf{m}_B) + 2 \times 0 + \text{Cov}(\mathbf{e}_A, \mathbf{e}_B) \\ &= \sigma_m^2 \mathbf{F}_A \mathbf{S}_{m,AB} \mathbf{F}_B^T + \sigma_e^2 \mathbf{S}_{e,AB}. \end{aligned} \tag{B6}$$

Thus, the joint distribution of the four  $\mathbf{m}_A$ ,  $\mathbf{m}_B$ ,  $\mathbf{d}_A$  and  $\mathbf{d}_B$ , conditioned on  $\sigma_e^2$  and  $\sigma_m^2$ , becomes

$$\begin{aligned} \begin{bmatrix} \mathbf{m}_A \\ \mathbf{m}_B \\ \mathbf{d}_A \\ \mathbf{d}_B \end{bmatrix} &= \mathcal{N} \left( \begin{bmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \\ \mathbf{F}_A \boldsymbol{\mu}_A \\ \mathbf{F}_B \boldsymbol{\mu}_B \end{bmatrix}, \begin{bmatrix} \sigma_m^2 \mathbf{S}_{m,AA} & & & \\ \sigma_m^2 \mathbf{S}_{m,BA} & & & \\ \sigma_m^2 \mathbf{F}_A \mathbf{S}_{m,AA} & & & \\ \sigma_m^2 \mathbf{F}_B \mathbf{S}_{m,BA} & & & \\ \sigma_m^2 \mathbf{S}_{m,AB} & \sigma_m^2 \mathbf{S}_{m,AA} \mathbf{F}_A^T & \sigma_m^2 \mathbf{S}_{m,AB} \mathbf{F}_B^T & \\ \sigma_m^2 \mathbf{S}_{m,BB} & \sigma_m^2 \mathbf{S}_{m,BA} \mathbf{F}_A^T & \sigma_m^2 \mathbf{S}_{m,BB} \mathbf{F}_B^T & \\ \sigma_m^2 \mathbf{F}_A \mathbf{S}_{m,AB} & \sigma_m^2 \mathbf{F}_A \mathbf{S}_{m,AA} \mathbf{F}_A^T + \sigma_e^2 \mathbf{S}_{e,AA} & \sigma_m^2 \mathbf{F}_A \mathbf{S}_{m,AB} \mathbf{F}_B^T + \sigma_e^2 \mathbf{S}_{e,AB} & \\ \sigma_m^2 \mathbf{F}_B \mathbf{S}_{m,BB} & \sigma_m^2 \mathbf{F}_B \mathbf{S}_{m,BA} \mathbf{F}_A^T + \sigma_e^2 \mathbf{S}_{e,BA} & \sigma_m^2 \mathbf{F}_B \mathbf{S}_{m,BB} \mathbf{F}_B^T + \sigma_e^2 \mathbf{S}_{e,BB} & \end{bmatrix} \right). \end{aligned} \tag{B7}$$

Grouping  $\mathbf{m}_B$ ,  $\mathbf{d}_A$  and  $\mathbf{d}_B$  into a vector denoted  $\mathbf{x}_R$  and correspondingly grouping the elements of the mean vector and covariance matrix, the joint distribution of  $\mathbf{m}_A$ ,  $\mathbf{m}_B$ ,  $\mathbf{d}_A$  and  $\mathbf{d}_B$  can be expressed as

$$\begin{bmatrix} \mathbf{m}_A \\ \mathbf{x}_R \end{bmatrix} = \mathcal{N} \left( \begin{bmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_R \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AR} \\ \boldsymbol{\Sigma}_{RA} & \boldsymbol{\Sigma}_{RR} \end{bmatrix} \right). \tag{B8}$$

Using the expression for conditional normal distribution, equation (C3), the proposal distribution thereby becomes  $\pi^{lin}(\mathbf{m}_A | \mathbf{m}_B, \mathbf{d}_A, \mathbf{d}_B, \sigma_m^2, \sigma_e^2) = \mathcal{N}(\mathbf{m}_A; \boldsymbol{\mu}_u, \boldsymbol{\Sigma}_u)$  with

$$\begin{aligned}\boldsymbol{\mu}_u &= \boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{AR} \boldsymbol{\Sigma}_{RR}^{-1} (\mathbf{x}_R - \boldsymbol{\mu}_R), \\ \boldsymbol{\Sigma}_u &= \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AR} \boldsymbol{\Sigma}_{RR}^{-1} \boldsymbol{\Sigma}_{RA}.\end{aligned}\tag{B9}$$

### APPENDIX C: STATISTICAL DISTRIBUTIONS

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

$$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\},\tag{C1}$$

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

For two jointly Gaussian variables  $\mathbf{x}_1$  and  $\mathbf{x}_2$ ,

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right),\tag{C2}$$

the conditional distribution for  $\mathbf{x}_1$  given  $\mathbf{x}_2$  is Gaussian with mean vector and covariance matrix given by

$$\begin{aligned}\boldsymbol{\mu}_{1|2} &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ \boldsymbol{\Sigma}_{1|2} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}.\end{aligned}\tag{C3}$$

The inverse gamma probability density function is

$$\mathcal{IG}(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left( \frac{1}{x} \right)^{\alpha+1} e^{-\beta/x},\tag{C4}$$

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

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

distribution of  $\sigma^2$  is

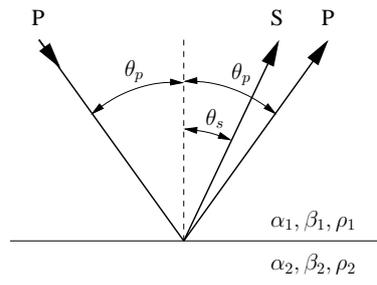
$$\begin{aligned}
 \pi(\sigma^2|\mathbf{x}) &\propto \pi(\mathbf{x}|\sigma^2) \pi(\sigma^2) \\
 &= \mathcal{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{S}) \mathcal{IG}(\alpha, \beta) \\
 &\propto \frac{1}{(\sigma^2)^{n/2} |\mathbf{S}|^{1/2}} \exp \left\{ -\frac{1}{2} \sigma^{-2} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{S}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\} \\
 &\quad \times \left( \frac{1}{\sigma^2} \right)^{\alpha+1} \exp \left\{ -\frac{\beta}{\sigma^2} \right\} \\
 &\propto \left( \frac{1}{\sigma^2} \right)^{\alpha+1+n/2} \exp \left\{ -\sigma^{-2} \left( \beta + s^2 \frac{n}{2} \right) \right\} \\
 &\propto \mathcal{IG} \left( \sigma^2 \mid \alpha + \frac{n}{2}, \beta + s^2 \frac{n}{2} \right),
 \end{aligned} \tag{C5}$$

where

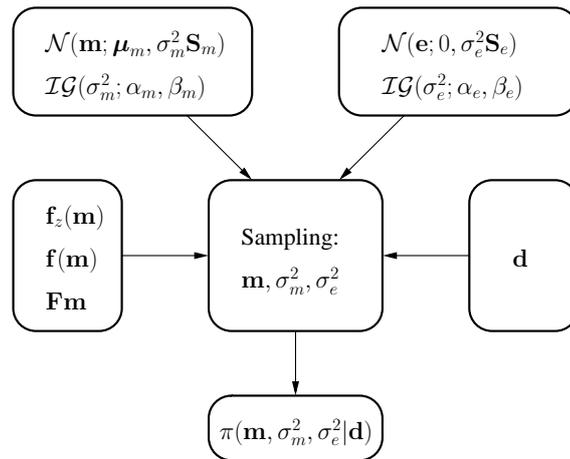
$$s^2 = \frac{1}{n} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{S}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \tag{C6}$$

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

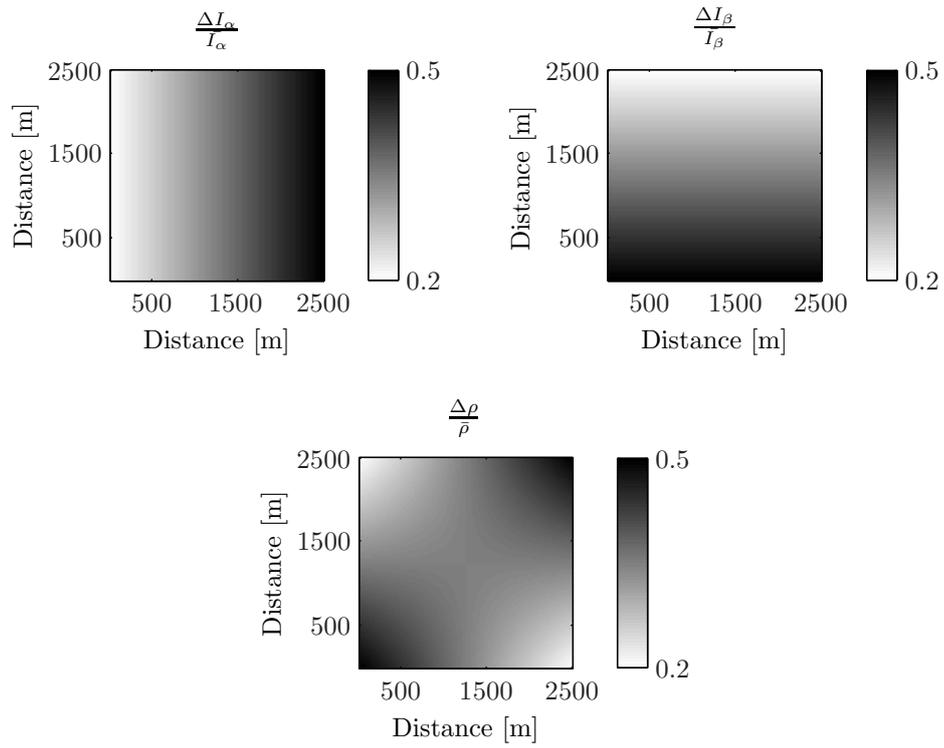
This paper has been produced using the Blackwell Publishing GJI L<sup>A</sup>T<sub>E</sub>X2e class file.



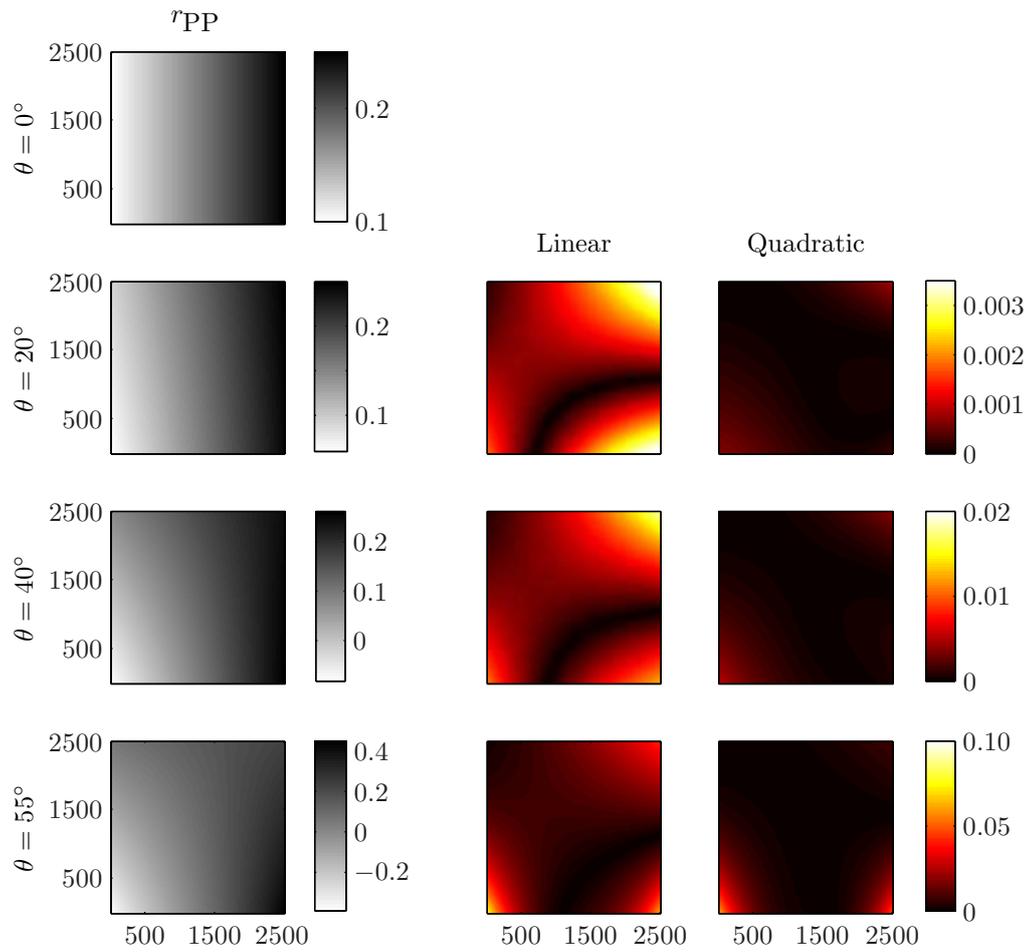
**Figure 1.** Directions of reflected P and S-waves for an incoming P-wave.



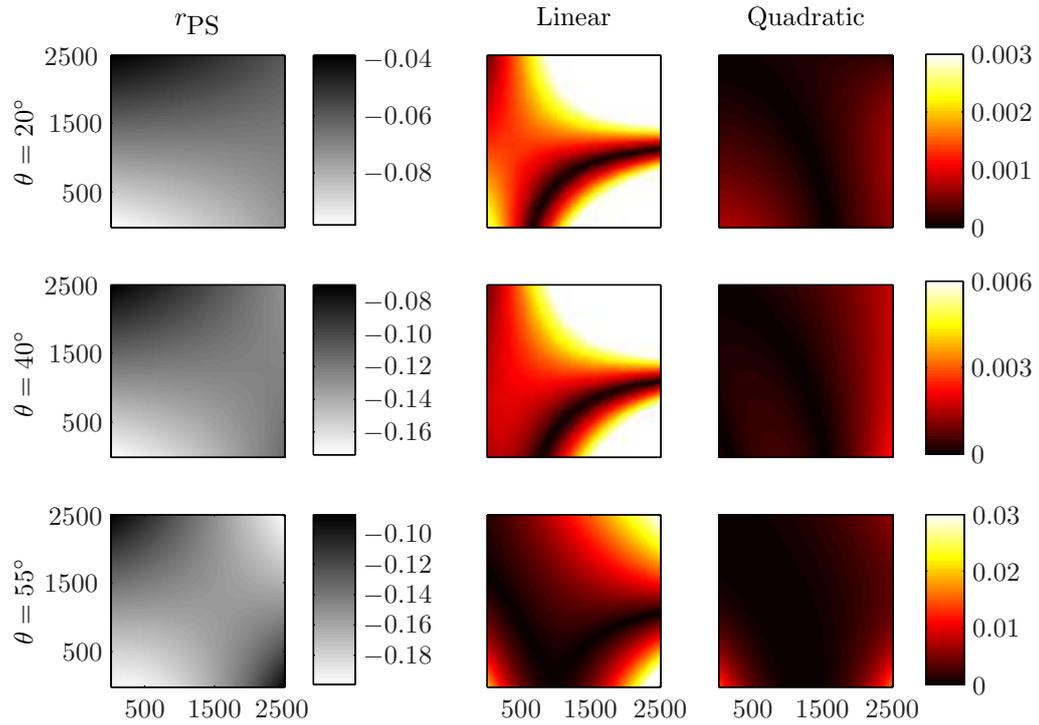
**Figure 2.** Input and output from the inversion algorithm. The sampling procedure is described in Section 3.



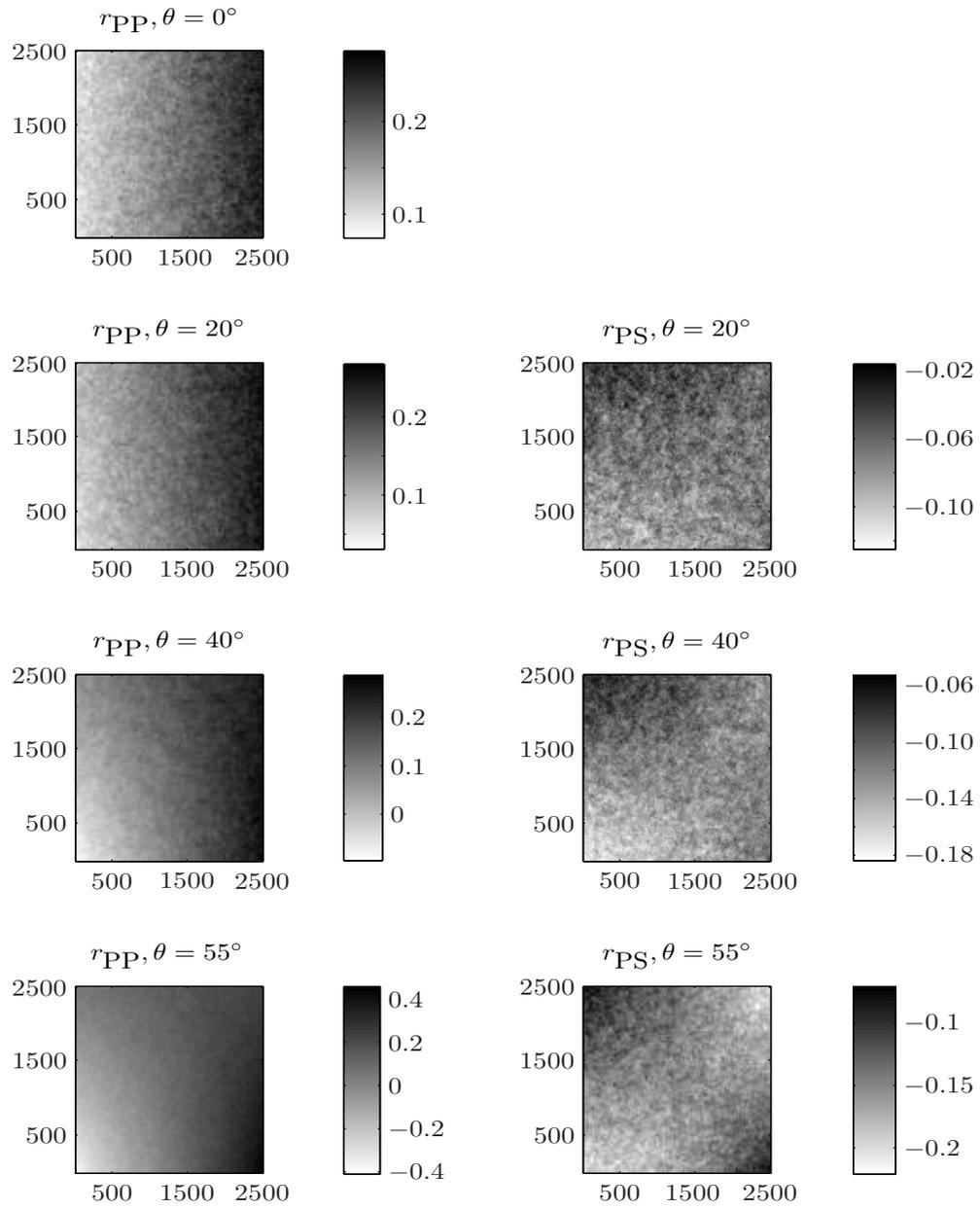
**Figure 3.** The true contrasts in  $\mathbf{m}$ . Upper left is P-wave impedance, upper right is S-wave impedance, and lower is density.



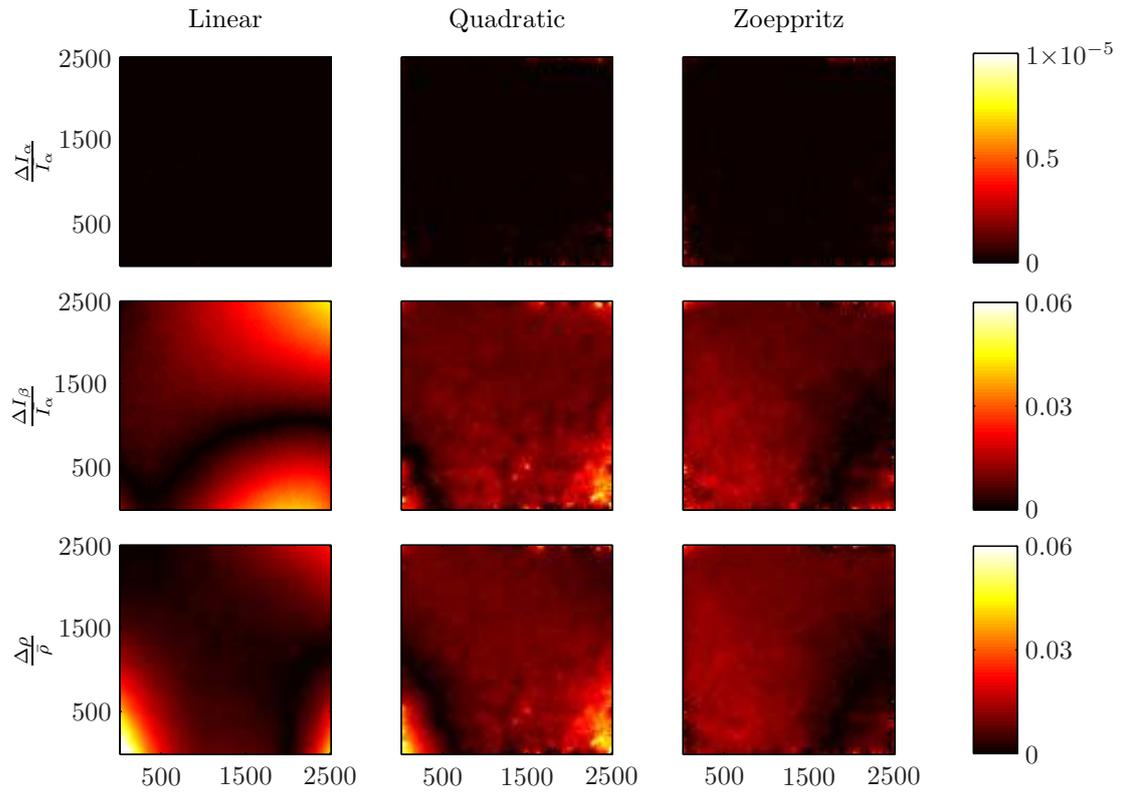
**Figure 4.** To the left is PP reflection coefficients from the Zoeppritz model for 4 different incidence angles. The two right columns show the bias in the linear and quadratic approximations, relative to the Zoeppritz model, for the nonzero angles. For  $\theta = 0^\circ$  the bias is zero.



**Figure 5.** 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 6.** PP and PS reflection coefficients from the Zoeppritz model with multivariate normally distributed noise.



**Figure 7.** 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.

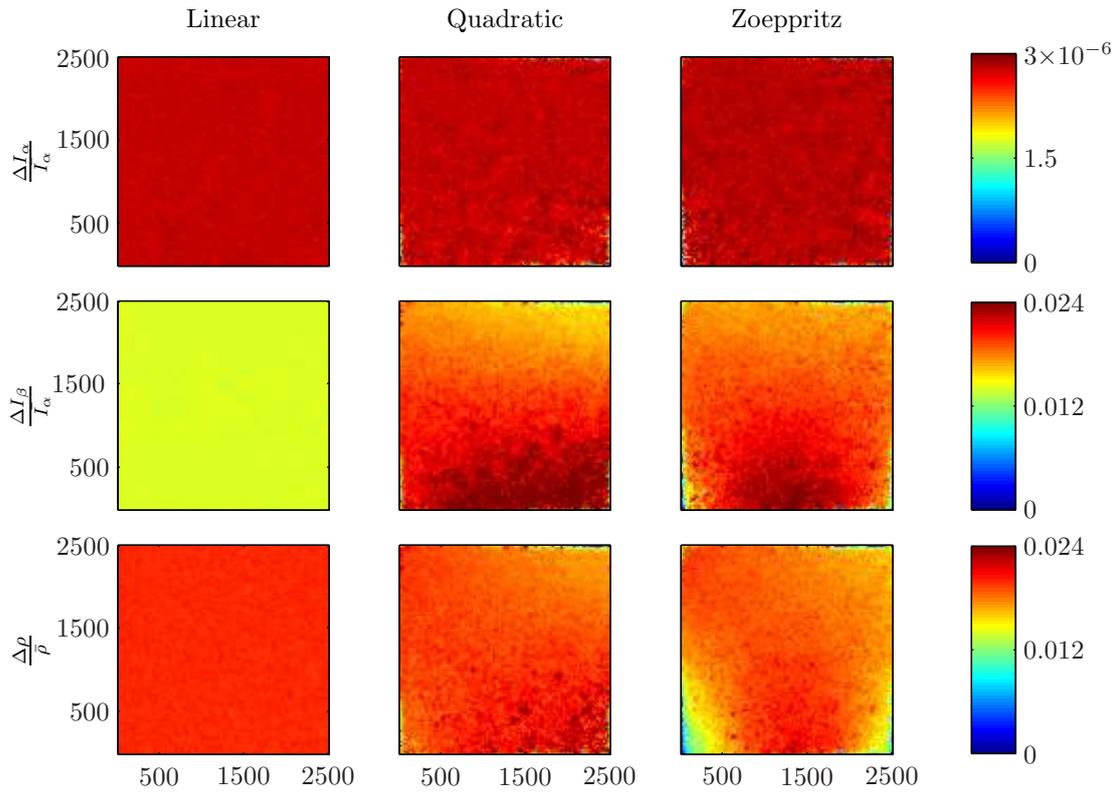
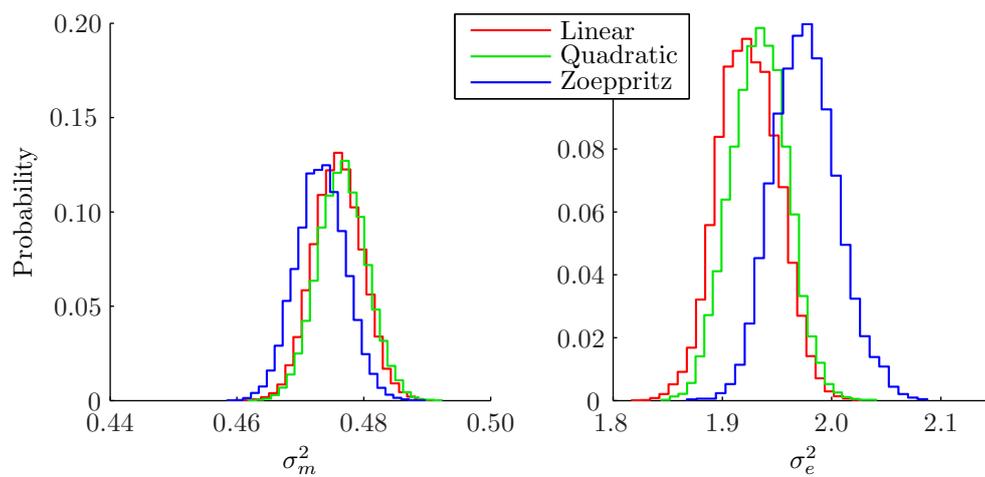
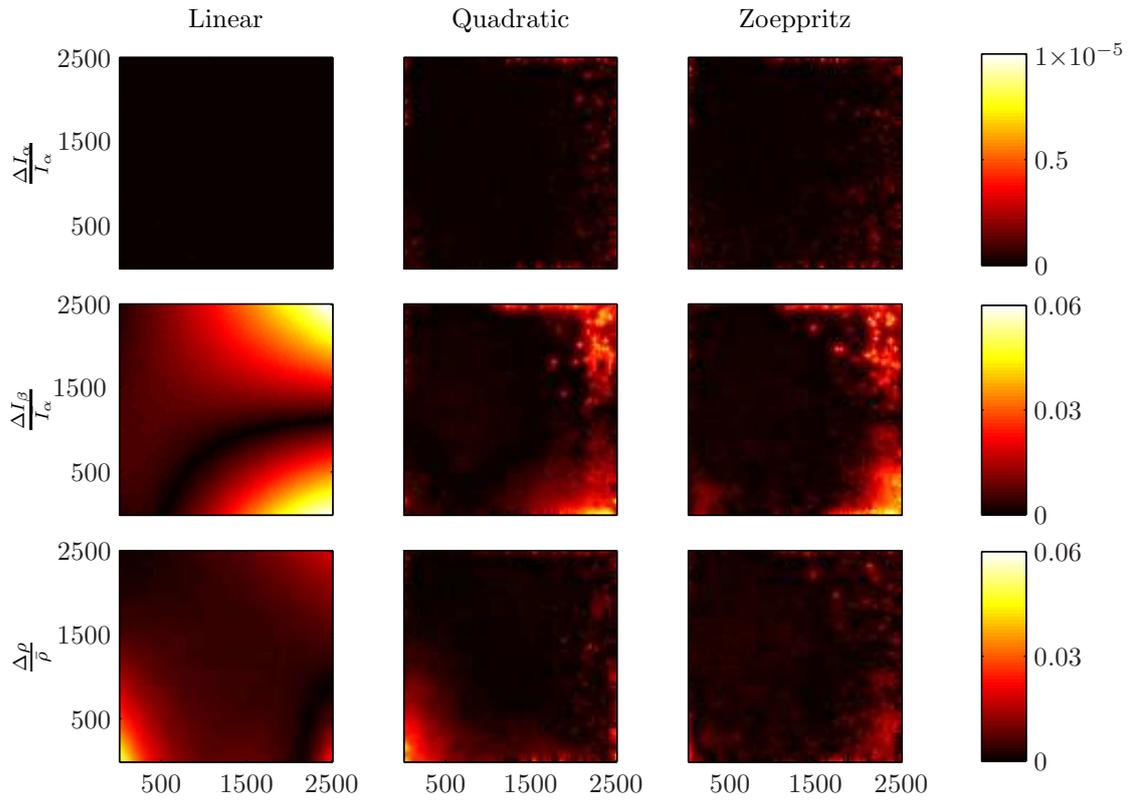


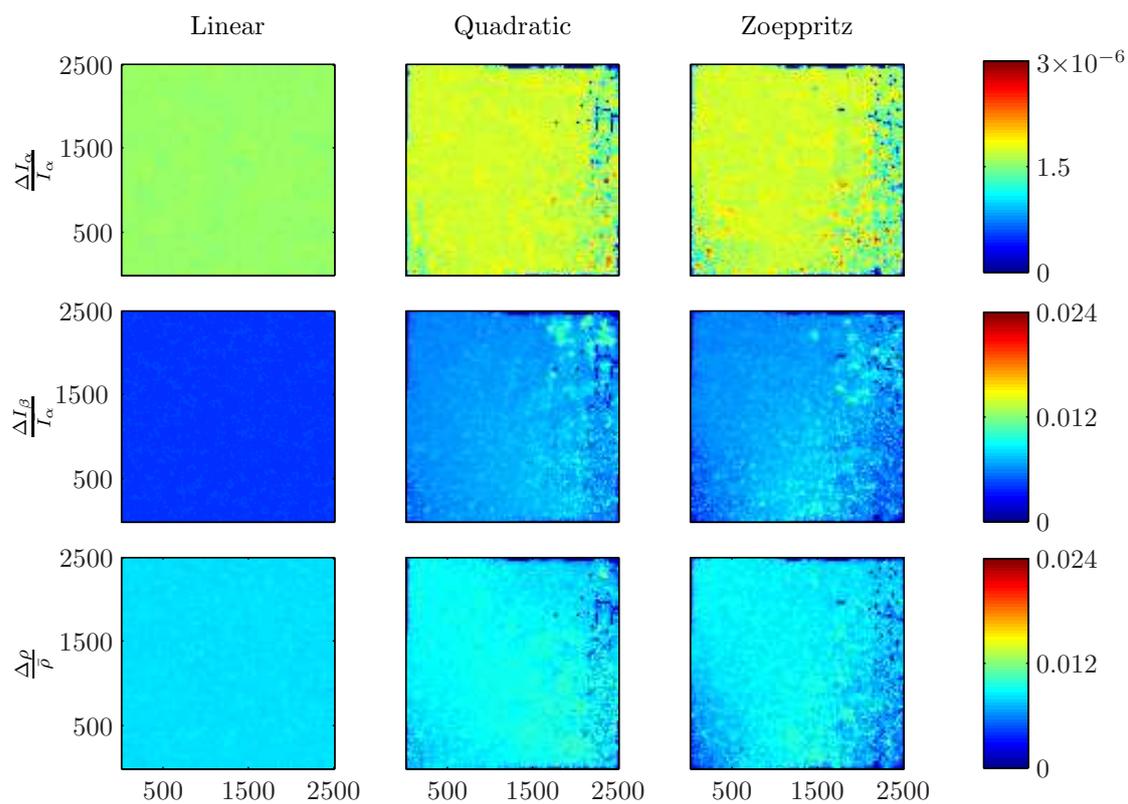
Figure 8. Standard deviation in the posterior distribution of  $\mathbf{m}$  from PP inversion without noise.



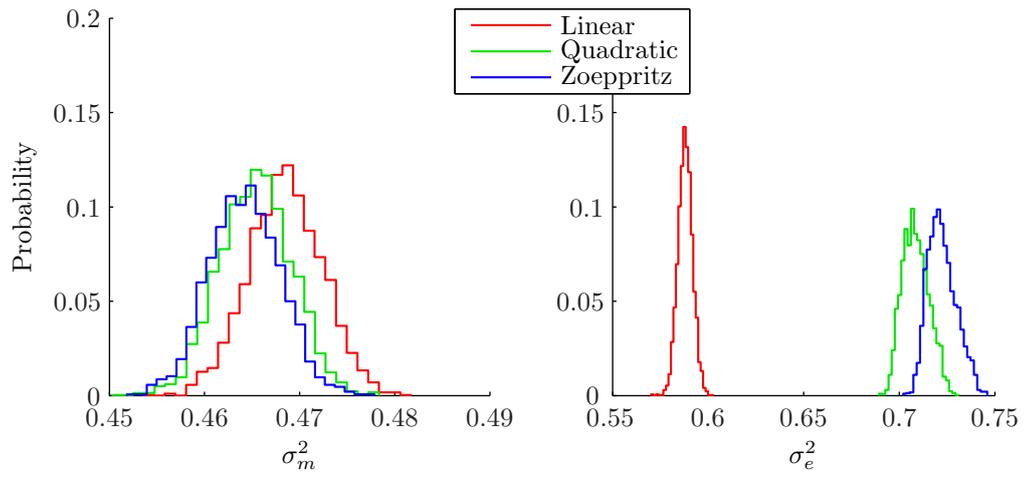
**Figure 9.** Posterior distribution of  $\sigma_m^2$  and  $\sigma_e^2$  from PP inversion without noise.



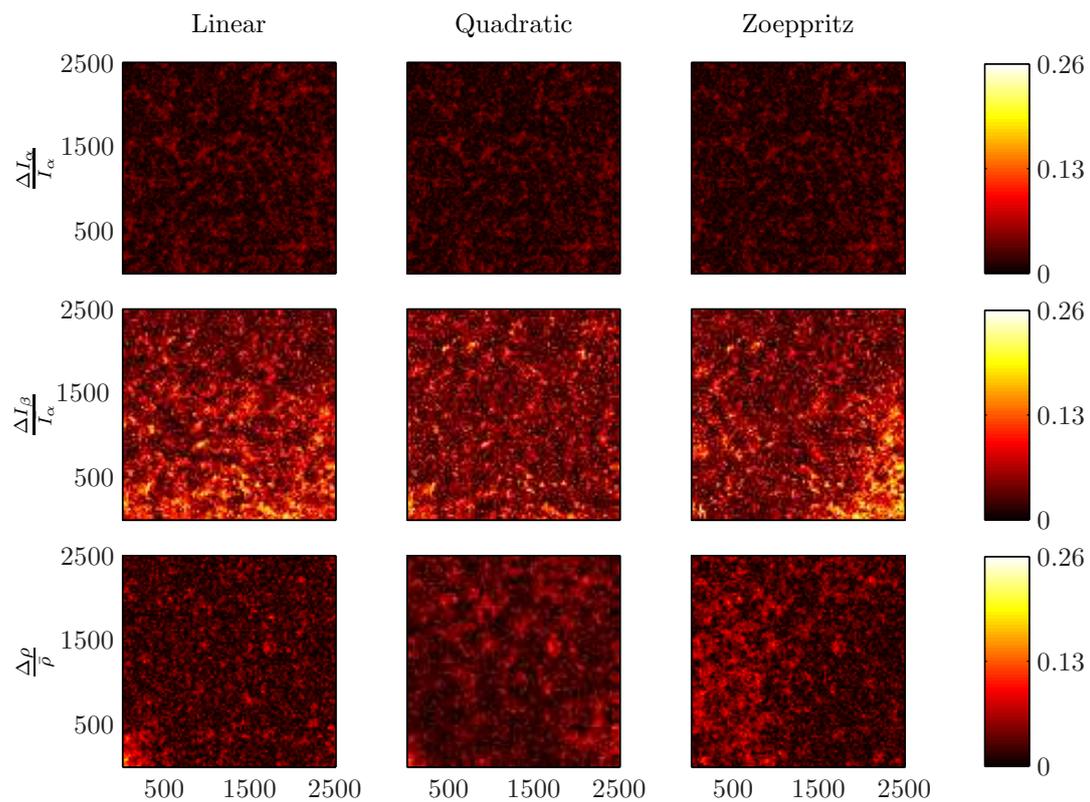
**Figure 10.** Bias in the posterior distribution of  $\mathbf{m}$  from joint PP and PS inversion without noise.



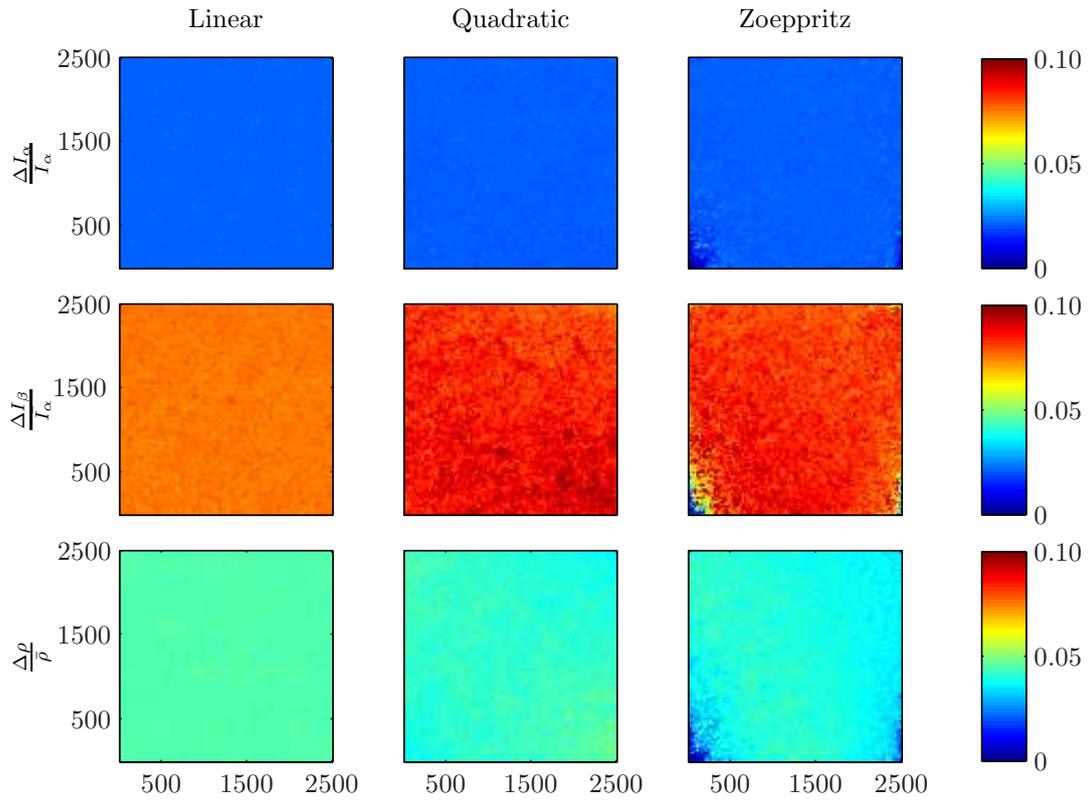
**Figure 11.** Standard deviation in the posterior distribution of  $\mathbf{m}$  from joint PP and PS inversion without noise.



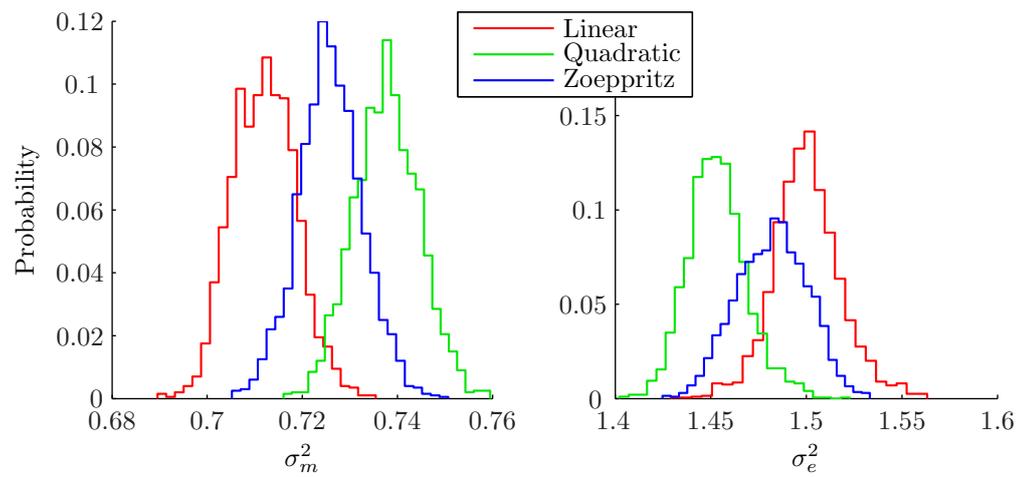
**Figure 12.** Posterior distribution of  $\sigma_m^2$  and  $\sigma_e^2$  from joint PP and PS inversion without noise.



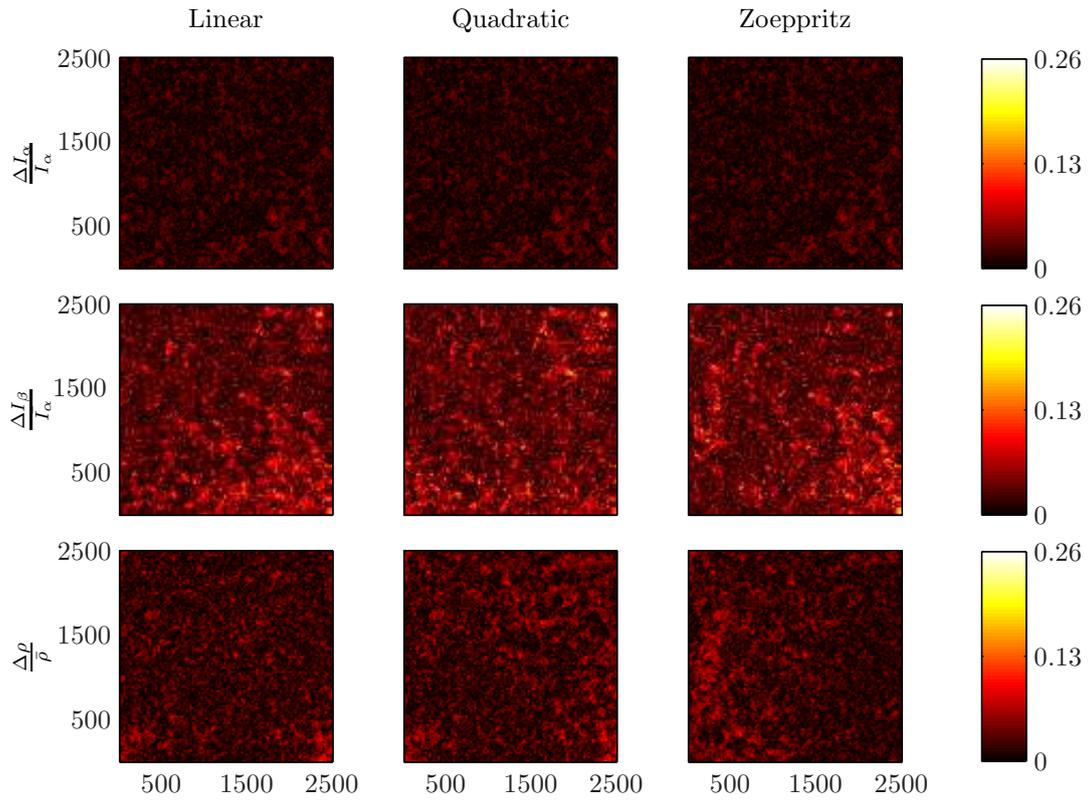
**Figure 13.** Bias in the posterior distribution of  $\mathbf{m}$  from PP inversion including multivariate Gaussian distributed noise.



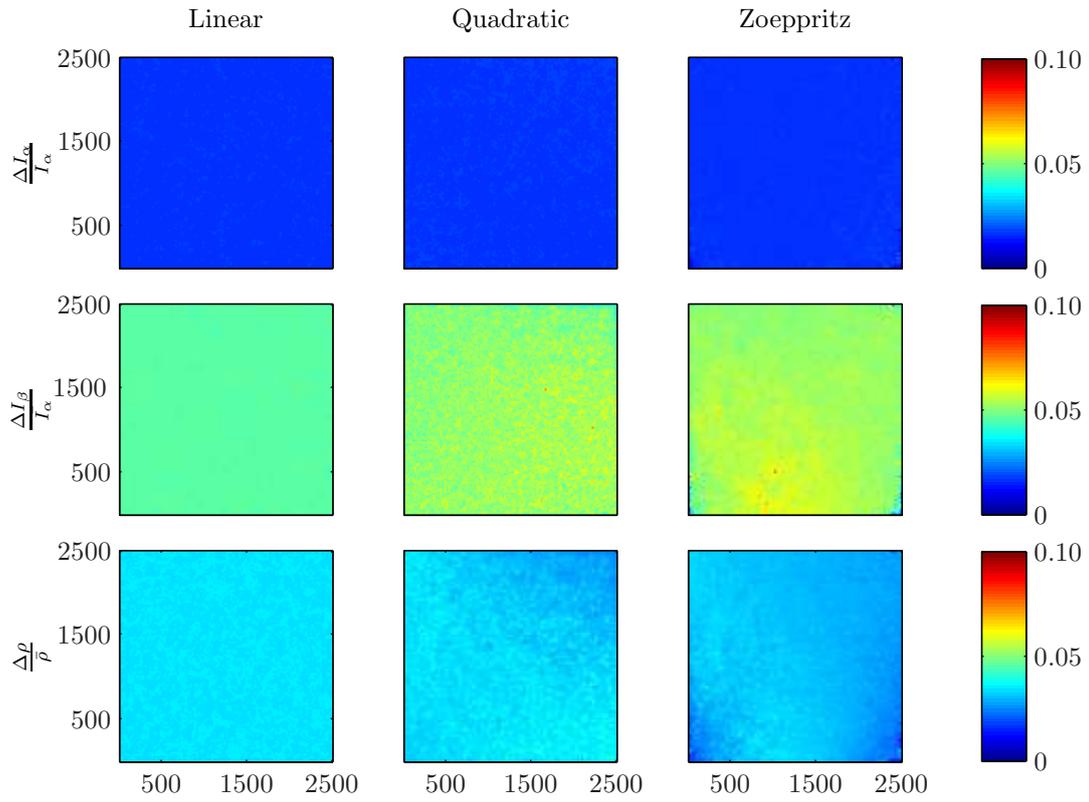
**Figure 14.** Standard deviation in the posterior distribution of  $\mathbf{m}$  from PP inversion including multivariate Gaussian distributed noise.



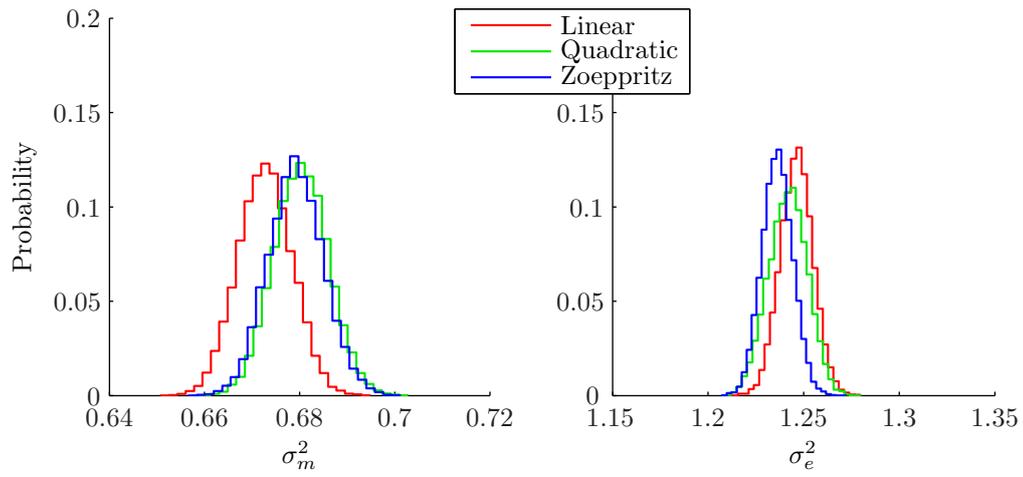
**Figure 15.** Posterior distribution of  $\sigma_m^2$  and  $\sigma_e^2$  from PP inversion including multivariate Gaussian distributed noise.



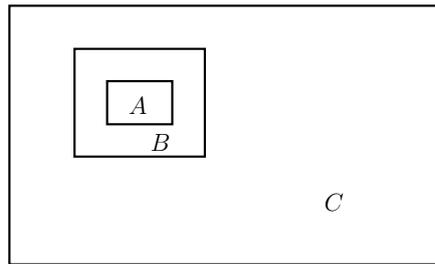
**Figure 16.** Bias in the posterior distribution of  $\mathbf{m}$  from joint PP and PS inversion including multivariate Gaussian distributed noise.



**Figure 17.** Standard deviation in the posterior distribution of  $\mathbf{m}$  from joint PP and PS inversion including multivariate Gaussian distributed noise.



**Figure 18.** Posterior distribution of  $\sigma_m^2$  and  $\sigma_e^2$  from joint PP and PS inversion including multivariate Gaussian distributed noise.



**Figure 19.** Block  $A$  is the block for which new values for  $\mathbf{m}$  are proposed,  $B$  is a boundary zone of limited thickness, and  $C$  is the rest.