Nonlinear least squares inversion of reflection coefficients using Bayesian regularization

Tor Erik Rabben and Bjørn Ursin, Norwegian University of Science and Technology*

SUMMARY

Inversion of seismic reflection coefficients is formulated in Bayesian framework. In addition to the elastic parameters two scalar quantities quantifying the variance level in the prior of the elastic parameters and measurement noise are included in the inversion. The maximum a posteriori solution is derived and the result is an adaptive weighted least squares inversion algorithm with the ratio between the variance levels as a data driven damping factor. The algorithm is tested on synthetic nonlinear PP and joint PP and PS reflection data, but is valid for inversion of any forward problem.

INTRODUCTION

Solving nonlinear least squares problems often involve ill-posed Hessian matrices, and in order to produce reliable results some kind of regularization is needed. We present here a least squares inversion algorithm which has a data driven regularization. The starting point is to formulate the inversion in a Bayesian framework, hence the name Bayesian regularization, where we include two scalars as inversion parameters. In our normal distributed prior and likelihood distributions they are variance level factors, see Buland and Omre (2003). Another possible solution is to explore the posterior distribution by a Metropolis-Hastings algorithm, see e.g. Tjelmeland and Eidsvik (2005), but here we only focus on the maximum a posteriori solution.

MODEL

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

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

and

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

where γ is the background v_S/v_P-ratio, θ_p the angle of the incoming and reflected P-wave, and θ_s the angle of the reflected S-wave. The parameters are collected in $\mathbf{m}$ and the measured reflection amplitudes in $\mathbf{d}$, both defined over a two dimensional grid

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

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

such that the total number of elements are $n_m = n_x n_y D_m$ and $n_e = n_x n_y D_d$. The forward model is the link from $\mathbf{m}$ to $\mathbf{d}$. Our focus will be on the quadratic approximations written

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

where $\mathbf{f}$ is (1) and (2) in case of both PP and PS reflections. We begin by assigning prior distributions to the model parameters $\mathbf{m}$ and the noise $\mathbf{e}$,

$$\pi(\mathbf{e}|\sigma_e^2) = N(\mathbf{e}; 0, \sigma_e^2\Sigma_e) \tag{6}$$

$$\pi(\mathbf{m}|\sigma_m^2) = N(\mathbf{m}; \mu_m, \sigma_m^2\Sigma_m). \tag{7}$$

Finding proper covariance matrices is difficult, we therefor include the estimation of the scalars σ_e^2 and σ_m^2 as a part of the inversion procedure and only specify the structure of the covariance. As a consequence, we need a prior distributions and choose the inverse gamma distribution

$$\pi(\sigma_e^2) = IG(\sigma_e^2; \alpha_e, \beta_e) \tag{8}$$

$$\pi(\sigma_m^2) = IG(\sigma_m^2; \alpha_m, \beta_m) \tag{9}$$

where α and β are scalar parameters defining the prior distributions, see e.g. Buland and Omre (2003) for the definition of the distribution. This distribution is a reasonable choice since it is defined only for positive values and has zero probability in the origin and infinity. From these four prior distributions and the forward model we can find the likelihood and, using Bayes rule, the joint posterior distribution.

MAXIMUM A POSTERIORI SOLUTION

Instead of trying to estimate the joint posterior $\pi(\mathbf{m}, \sigma_e^2, \sigma_m^2|\mathbf{d})$ we will update each parameter sequentially in an iterative algorithm. We therefore need the three posterior expressions

$$\pi(\mathbf{m}|\mathbf{d}, \sigma_e^2, \sigma_m^2) \propto \pi(\mathbf{d}|\mathbf{m}, \sigma_e^2)\,\pi(\mathbf{m}|\sigma_m^2) \tag{10}$$

$$\pi(\sigma_e^2|\mathbf{d}, \mathbf{m}) \propto \pi(\mathbf{d}|\mathbf{m}, \sigma_e^2)\,\pi(\sigma_e^2) \tag{11}$$

$$\pi(\sigma_m^2|\mathbf{m}) \propto \pi(\mathbf{m}|\sigma_m^2)\,\pi(\sigma_m^2). \tag{12}$$

Each of these could be sampled and hence assessing both mean and uncertainties, but a much faster algorithm will be to only search for the most likely solution, also known as the maximum a posteriori (MAP) solution. For the posterior (10) we write

$$\begin{aligned} \pi(\mathbf{m}|\mathbf{d}, \sigma_e^2, \sigma_m^2) \propto & \exp\left\{-\frac{1}{2\sigma_e^2}[\mathbf{d}-\mathbf{f}(\mathbf{m})]^T\Sigma_e^{-1}[\mathbf{d}-\mathbf{f}(\mathbf{m})]\right\} \\ & \times\exp\left\{-\frac{1}{2\sigma_m^2}[\mathbf{m}-\mu_m]^T\Sigma_m^{-1}[\mathbf{m}-\mu_m]\right\} \end{aligned} \tag{13}$$

Maximizing the posterior is equal to minimizing the expression

$$\varphi = \frac{1}{2\sigma_e^2}||\mathbf{d}-\mathbf{f}(\mathbf{m})||^2_{\Sigma_e^{-1}} + \frac{1}{2\sigma_m^2}||\mathbf{m}-\mu_m||^2_{\Sigma_m^{-1}}, \tag{14}$$

which is a weighted least squares problem. The minimum is reached when the gradient of φ is zero, and by expanding it in a Taylor series we find the iterative solution algorithm

$$\begin{aligned}\mathbf{m}_{k+1} &= \mathbf{m}_k - \mathbf{H}_k^{-1}\mathbf{g}_k \\ &= \mathbf{m}_k - \left(\mathbf{J}^T\Sigma_e^{-1}\mathbf{J} + \lambda^2\Sigma_m^{-1}\right)^{-1}\left(\lambda^2\Sigma_m^{-1}\Delta\mathbf{m}_0 - \mathbf{J}^T\Sigma_e^{-1}\Delta\mathbf{d}\right)\end{aligned} \tag{15}$$

where $\mathbf{J} = \partial\mathbf{f}/\partial\mathbf{m}^T$, $\Delta\mathbf{d} = \mathbf{d} - \mathbf{f}(\mathbf{m}_k)$, $\Delta\mathbf{m}_0 = \mathbf{m}_k - \mu_m$, and $\lambda^2 = \sigma_e^2/\sigma_m^2$. By skipping the first term in the gradient and assuming $\Sigma_e = \Sigma_m = \mathbf{I}$ it reduces to the famous Levenberg-Marquardt algorithm, see e.g. Tarantola (1987).

The posterior (11) and (12) can, by using the definition of the normal and inverse gamma distribution, be written

$$\pi(\sigma_e^2|\mathbf{d},\mathbf{m}) \propto IG\left(\sigma_e^2; \alpha_e + \frac{n_e}{2}, \beta_e + \frac{s_e^2}{2}\right) \tag{16}$$

$$\pi(\sigma_m^2|\mathbf{m}) \propto IG\left(\sigma_m^2; \alpha_m + \frac{n_m}{2}, \beta_m + \frac{s_m^2}{2}\right) \tag{17}$$

with $s_e^2 = ||\Delta\mathbf{d}||^2_{\Sigma_e^{-1}}$ and $s_m^2 = ||\Delta\mathbf{m}_0||^2_{\Sigma_m^{-1}}$. The MAP solution of an inverse gamma distribution $IG(\sigma^2;\alpha,\beta)$ is $\sigma^2 = \beta/(\alpha+1)$, and with this result we find the MAP of (16) and (17).

To speed up the algorithm we will perform only one update of $\mathbf{m}$ using (15) before updating the MAP of σ_e^2 and σ_m^2. The final expression for λ^2 reads

$$\lambda^2_{k+1} = \frac{\sigma^2_{e,k+1}}{\sigma^2_{m,k+1}} = \frac{\beta_e + \frac{1}{2}s^2_{e,k+1}}{\beta_m + \frac{1}{2}s^2_{m,k+1}} \cdot \frac{1 + \alpha_m + \frac{1}{2}n_m}{1 + \alpha_e + \frac{1}{2}n_e}. \tag{18}$$

Equations (15) and (18) constitute the basis of our inversion algorithm.

IMPLEMENTATION

For the covariance matrices in (6) and (7) we choose to split them in two parts

$$\Sigma_e = g_e \otimes S_e \tag{19}$$

$$\Sigma_m = g_m \otimes S_m. \tag{20}$$

Here, g_e and g_m are matrices describing covariances between parameters and errors within each spatial point, and S_e and S_m are spatial correlation matrices. The symbol $\otimes$ denotes the Kronecker product. The dimensions of g_e and g_m are $D_d \times D_d$ and $D_m \times D_m$ respectively. To generate the correlation matrices we use the exponential correlation function

$$S(i,j) = \exp\left\{-3\frac{|x_i - x_j|}{d}\right\} \tag{21}$$

where d is known as the range of the correlation. The length between the points x_i and x_j, $|x_i - x_j|$, is defined on a torus. This will produce edge effects, but they are limited to a distance proportional to the range parameter. The advantage of this assumption is that the matrices S_e and S_m become circulant which in turn enables fast calculation of the inverse. For details on solving circulant matrices see Rue and Held (2005).

To solve (15) is the major part of the computational time. The Hessian matrix $\mathbf{H}$ can easily become too large to even be stored in memory. Our solution is to use an iterative solver and because the Hessian is symmetric positive definite (SPD) we will use the conjugate gradient method (CG) to solve the system

$$\mathbf{H}_k\Delta\mathbf{m} = \mathbf{g}_k, \tag{22}$$

for $\Delta\mathbf{m}$ where $\Delta\mathbf{m} = \mathbf{m}_{k+1} - \mathbf{m}_k$ (not to be confused with $\Delta\mathbf{m}_0 = \mathbf{m} - \mu_m$). In CG the most computational expensive is the the matrix-vector product of the type $\mathbf{H}_k\mathbf{p}$, but can be evaluated very fast by exploiting the sparse structure of the Jacobian $\mathbf{J}$ and that S is circulant. In addition, to improve the convergence properties, we have used the diagonal of $\mathbf{H}_k$, D, as a precondition matrix and solved the preconditioned system

$$L^{-1}\mathbf{H}_kL^{-T}u = L^{-1}\mathbf{g}_k, \quad \Delta\mathbf{m} = L^{-T}u, \tag{23}$$

where $D = LL^T$, the Cholesky decomposition. For details on preconditioning and CG see Saad (2000).

NUMERICAL EXAMPLE

We have used a synthetic model to test the inversion algorithm. From a chosen true $\mathbf{m}$ we used the exact Zoeppritz equations to generate synthetic measurements $\mathbf{d}$ by assuming the P-wave velocity in the upper medium and the background v_P/v_S ratio to be known. Our true $\mathbf{m}$ was defined on a 100×100 grid, see (3), and ranging from 0.2 to 0.5. In Figures 1 and 2 we show the synthetic data $\mathbf{d}$ together with the bias in the linear and quadratic approximations for PP and PS data respectively. It is clear that the quadratic approximations are superior. For the prior mean μ_m we have used half of the true value of $\mathbf{m}$, and this is also used as initial $\mathbf{m}_0$.

The first example is the inversion of four PP reflection amplitudes. In Figure 3(a) we see the development of λ^2 in each iteration showing a convergent behavior. In the first iteration we have limited the maximum value to be 5, otherwise it would have been several orders of magnitude higher because the initial guess $\mathbf{m}_0$ is far from the truth. The value in the final iteration is approximately 0.6. The absolute value of the bias of the corresponding $\mathbf{m}$ is displayed in Figure 4, using both the linear, quadratic, and exact Zoeppritz as forward model. We see that P-wave impedance is well defined in all three cases, while the two other parameters have bias for the two approximate models. The bias is clearly correlated with the bias shown in Figure 1.

The second example is very similar to the first, but this time we also include the three PS reflection amplitudes such that the total number of amplitudes, D_d, is equal to seven. λ_k^2 for each iteration is shown in Figure 3(b). Again it shows nice convergent behavior and in the final iteration it is 0.3. Compared to the previous example we see that the value is lower and the inversion problem is less damped, or, in other words, more weight is put on the measured data. This is also reflected in Figure 5 where the absolute value of the bias is plotted, although for the linear model there is still a considerable bias.

CONCLUSION

From a Bayesian formulation of the inversion problem we have derived a weighted least squares algorithm with an adaptive, data driven damping factor. The algorithm was tested on two synthetic examples and showed a well-posed behavior and a low bias in the inversion result.

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.

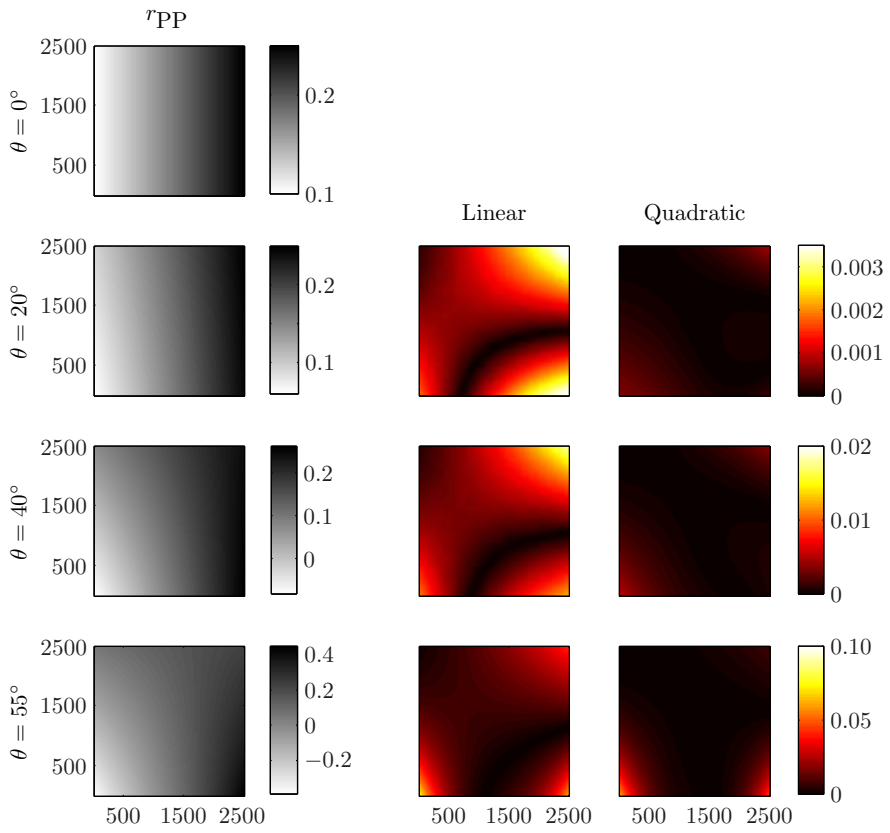

Figure 1: To the left is PP reflection coefficients from the Zoeppritz model for 4 different incidence angles. The two right columns show the bias in the linear and quadratic approximations, relative to the Zoeppritz model, for the nonzero angles. For $\theta = 0°$ the bias is zero.

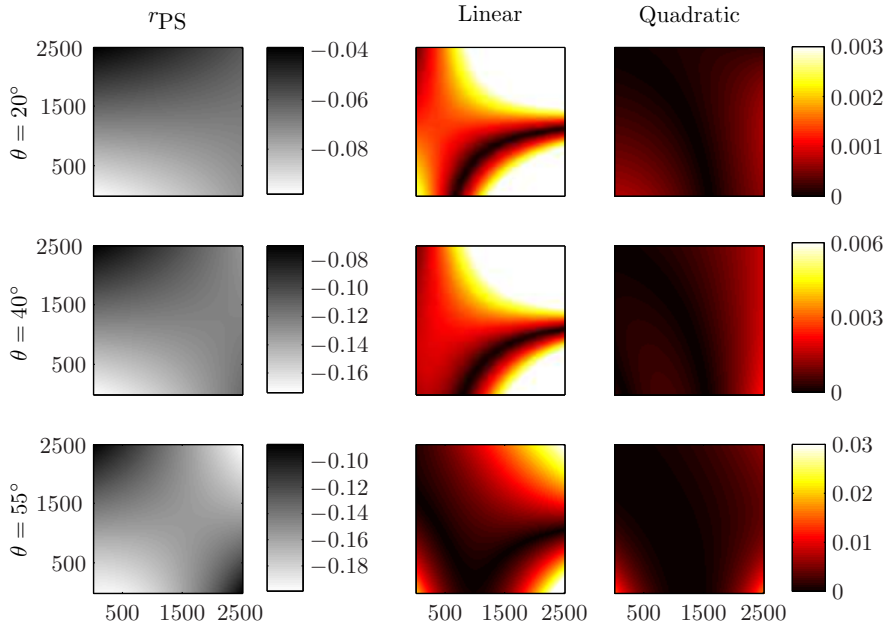

Figure 2: To the left is PS reflection coefficients from the Zoeppritz model for 3 nonzero incidence angles. The two right columns show the bias in the linear and quadratic approximations, relative to the Zoeppritz model.

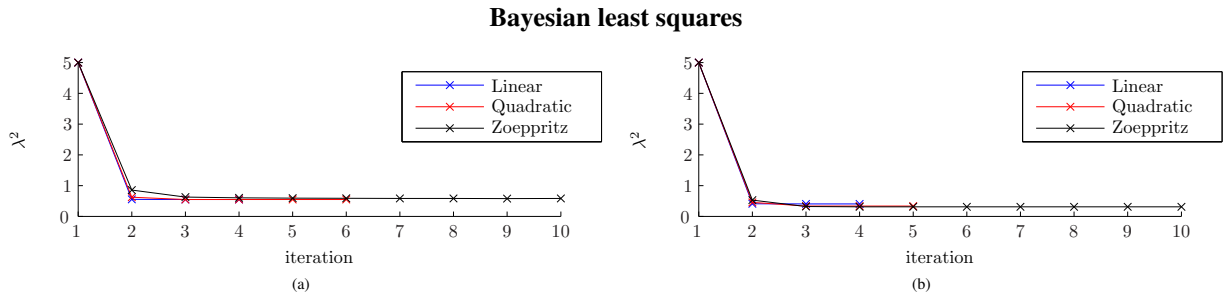

Figure 3: λ_k^2 for (a) PP inversion and (b) joint PP and PS inversion.

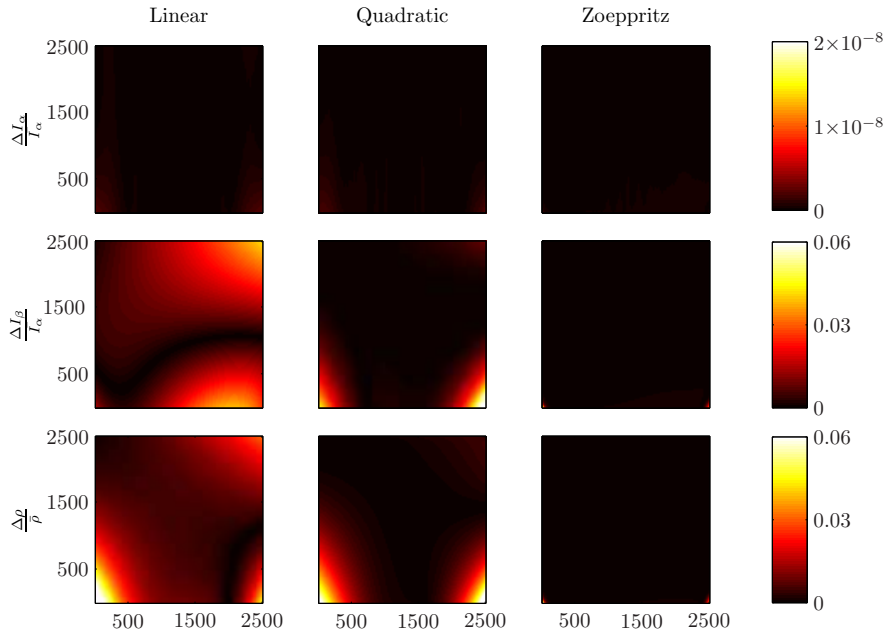

Figure 4: Absolute value of the bias in PP inversion.

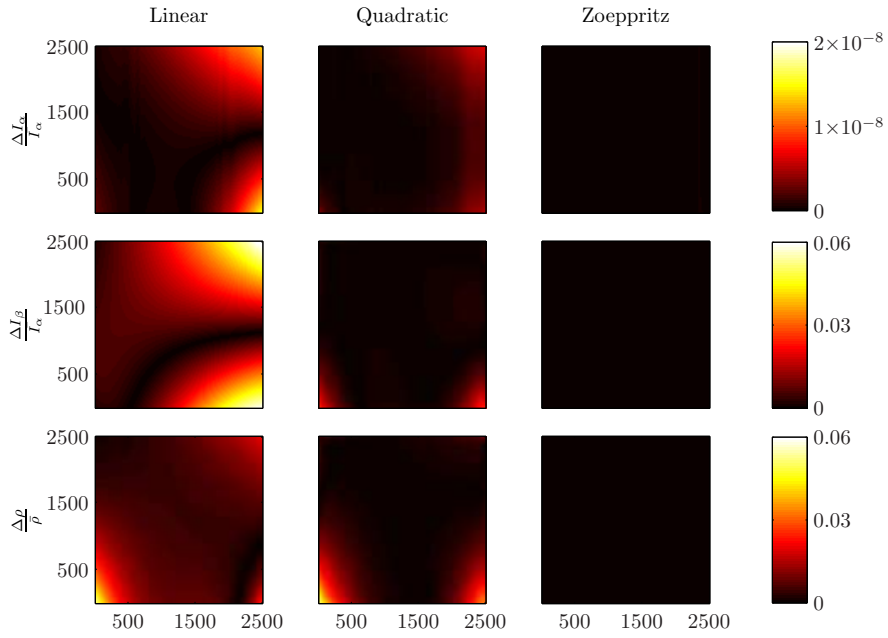

Figure 5: Absolute value of the bias in joint PP and PS inversion.

REFERENCES

Buland, A. and H. Omre, 2003, Joint AVO inversion, wavelet estimation and noise-level estimation using a spatially coupled hierarchical bayesian model: Geophysical Prospecting, **51**, 531–550.

Rue, H. and L. Held, 2005, Gaussian markov random fields: Chapman & Hall/CRC.

Saad, Y., 2000, Iterative methods for sparse linear systems: PWS.

Stovas, A. and B. Ursin, 2003, Reflection and transmission responses of layered transversely isotropic viscoelastic media: Geophysical Prospecting, **51**, 447–477.

Tarantola, A., 1987, Inverse problem theory: Elsevier.

Tjelmeland, H. and J. Eidsvik, 2005, Directional Metropolis-Hastings update for posteriors with non-linear likelihoods: Presented at the Geostatistics Banff 2004.