Simultaneous Estimation of Reflectivity and Geologic Texture: Least-Squares Migration with a Hierarchical Bayesian Model

S. Ahmad Zamanian, MIT Earth Resources Laboratory; Jonathan Kane, Shell International E&P, Inc.; William Rodi, Michael Fehler, MIT Earth Resources Laboratory*

SUMMARY

In many geophysical inverse problems, smoothness assumptions on the underlying geology are utilized to mitigate the effects of poor resolution and noise in the data and to improve the quality of the inferred model parameters. Within a Bayesian inference framework, *a priori* assumptions about the probabilistic structure of the model parameters impose such a smoothness constraint or regularization. We consider the particular problem of inverting seismic data for the subsurface reflectivity of a 2-D medium, where we assume a known velocity field. In particular, we consider a hierarchical Bayesian generalization of the Kirchhoff-based least-squares migration (LSM) problem. We present here a novel methodology for estimation of both the optimal image and regularization parameters in a least-squares migration setting. To do so we utilize a Bayesian statistical framework that treats both the regularization parameters and image parameters as random variables to be inferred from the data. Hence rather than fixing the regularization parameters prior to inverting for the image, we allow the data to dictate where to regularize. In order to construct our prior model of the subsurface and regularization parameters, we define an undirected graphical model (or Markov random field) where vertices represent reflectivity values, and edges between vertices model the degree of correlation (or lack thereof) between the vertices. Estimating optimal values for the vertex parameters gives us an image of the subsurface reflectivity, while estimating optimal edge strengths gives us information about the local "texture" of the image, which, in turn, may tell us something about the underlying geology. Subsequently incorporating this information in the final model produces more clearly visible discontinuities in the final image. The inference framework is verified on a 2-D synthetic dataset, where the hierarchical Bayesian imaging results significantly outperform standard LSM images.

INTRODUCTION

Seismic imaging (or migration) refers to the process of creating an image of the Earth's subsurface reflectivity from seismic data, consisting typically of a set of seismograms generated by sources and recorded by receivers located at or near the surface. We denote the seismic modeling process as $\mathbf{d} = A\mathbf{m}$, where $\mathbf{d}$ is the seismic data, A is the forward modeling operator, and $\mathbf{m}$ is the subsurface reflectivity model (or seismic image). Traditional migration methods for estimating the image typically involve operating on the seismic data with the adjoint of the forward modeling operator (Claerbout, 1992) possibly along with a modifying function which attempts to correct for amplitude loss due to geometric spreading, transmission, absorption, etc. (Bleistein, 1984; Hanitzsch et al., 1994). More recently, attempts have been made to solve the imaging problem when cast as a least-squares inverse problem (Nemeth et al., 1999; Duquet et al., 2000). This approach to imaging is conventionally referred to as *least-squares migration*. As a least-squares inverse problem, it is typical to impose a form of regularization in the LSM cost function (that would typically penalize less smooth images). From the perspective of Bayesian inference, the regularization defines a prior probability distribution on the image parameters. Importantly however, the Bayesian framework also provides a rigorous mathematical framework for going beyond simply fixing a regularization on the image parameters. Indeed, under a hierarchical Bayesian setting, one can assign a prior probability distribution *to the regularization parameters*, which can then be inferred from the data or integrated out, resulting in potentially better inference for the image parameters.

Early treatments of the seismic imaging problem as a least-squares inverse problem can be found in LeBras and Clayton (1988) and Lambare et al. (1992). Kirchhoff-based LSM uses a ray-theory based forward modeling operator; its derivation and application is discussed in Nemeth et al. (1999) and Duquet et al. (2000). LSM has also been applied to a wave-equation-based forward modeling operator, as shown by Kühl and Sacchi (2003). Clapp (2005) describes two methods for choosing a regularization scheme in the LSM context where the image is constrained to be smooth either along geological features predetermined by a seismic interpreter or along the ray-parameter axis. Regarding the application of hierarchical Bayesian approaches to other geophysical inverse problems, Malinverno and Briggs (2004) applied hierarchical and empirical Bayesian methods to a 1-D traveltime tomography problem. Malinverno (2002, 2000) also applied Bayesian model selection, implemented via 'reversible-jump' Markov chain Monte Carlo (rj-MCMC) (Green, 1995), to determine an appropriate model parameterization and dimension when inverting for 1-D density and resistivity models from gravity gradient measurements and DC resistivity sounding data, respectively. Buland and Omre (2003) considered a hierarchical Bayesian approach to amplitude versus offset (AVO) inversion for the velocity and density model of a 2-D medium, where the seismic source wavelet is also estimated from the data. In an example of an inverse problem on a continental scale, Bodin et al. (2012) applied trans-dimensional Bayesian model selection to determine group velocities for the Australian continent from Rayleigh wave group delays (again implemented via rj-MCMC).

METHODOLOGY

Standard Kirchhoff-based LSM Framework

We restrict our attention to Kirchhoff-based least-squares migration, which attempts to invert the forward problem $\mathbf{d} = A\mathbf{m}$, where A is the Kirchhoff modeling operator. In particular, least-squares migration seeks to find the image $\mathbf{m}_{LS}$ that minimizes the ℓ^2-norm of the residual (the difference between the observed data $\mathbf{d}$ and the modeled data $\hat{\mathbf{d}} = A\mathbf{m}$). Without regularization, the LSM image is given by

$$\mathbf{m}_{LS} = \arg\min_{\mathbf{m}} \|\mathbf{d} - A\mathbf{m}\|_2^2. \quad (1)$$

Regularization is often introduced by adding a term which penalizes differences between model parameters and an additional term which penalizes the magnitude of the image (primarily to ensure well-posedness) to the LSM cost function. This gives the regularized LSM image as

$$\mathbf{m}_{LS} = \arg\min_{\mathbf{m}} \|\mathbf{d} - A\mathbf{m}\|_2^2 + \lambda \sum_{(i,j)\in E} \beta_{ij}(m_i - m_j)^2 + \varepsilon \sum_i m_i^2 \quad (2)$$

$$= \arg\min_{\mathbf{m}} \|\mathbf{d} - A\mathbf{m}\|_2^2 + \mathbf{m}^T (\lambda D(\beta) + \varepsilon I)\,\mathbf{m}, \quad (3)$$

where $\beta_{ij} \in \{0,1\}$ indicates whether or not to penalize the difference between m_i and m_j, E is the set of all pairs of image parameter indices where we could potentially penalize differences, λ assigns the weight given to penalizing these differences, and ε is a typically very small number that assigns the weight given to penalizing the magnitude of the image. Equation 3 is simply Equation 2 rewritten in compact matrix-vector notation, where D is a differencing operator defined by the vector $\beta = \{\beta_{ij} : (i,j) \in E\}$. Taking the derivative of the right-hand side of Equation 3 and setting it to zero yields the solution to the regularized LSM problem:

$$\mathbf{m}_{LS} = \left(A^T A + \lambda D(\beta) + \varepsilon I\right)^{-1} A^T \mathbf{d}. \quad (4)$$

Note that the εI term ensures that $A^T A + \lambda D(\beta) + \varepsilon I$ is positive definite, so that the regularized least-squares solution is guaranteed to exist and be unique.

Bayesian Framework

Standard Bayesian Formulation

It is possible to retrieve the very same LSM result when formulating the imaging problem in a Bayesian framework, where the image $\mathbf{m}$ and the data $\mathbf{d}$ are modeled as random vectors. In particular, we model $\mathbf{m}$ *a priori* as being Gaussian with zero mean and some covariance matrix Λ (i.e. $\mathbf{m} \sim N(0,\Lambda)$). We again model the seismic data as before with A our Kirchhoff modeling operator and additionally add zero-mean Gaussian noise $\mathbf{n}$ with some covariance matrix Σ to the data so that $\mathbf{d} = A\mathbf{m} + \mathbf{n}$ where $\mathbf{n} \sim N(0,\Sigma)$. Thus the conditional distribution for the data $\mathbf{d}$ given the model $\mathbf{m}$ will be Gaussian so that $\mathbf{d}|\mathbf{m} \sim N(A\mathbf{m},\Sigma)$. Applying Bayes' rule reveals that the posterior distribution for the model $\mathbf{m}$ conditioned on the data $\mathbf{d}$ is also Gaussian

$$\mathbf{m}|\mathbf{d} \sim N(\mu_{\text{post}}, \Lambda_{\text{post}}) \quad (5)$$

with posterior mean

$$\mathbb{E}[\mathbf{m}|\mathbf{d}] = \mu_{\text{post}} = \left(A^T \Sigma^{-1} A + \Lambda^{-1}\right)^{-1} A^T \Sigma^{-1} \mathbf{d} \quad (6)$$

and posterior covariance matrix

$$\Lambda_{\text{post}} = \left(A^T \Sigma^{-1} A + \Lambda^{-1}\right)^{-1}. \quad (7)$$

The mean of the posterior distribution given in Equation 6 is what is referred to as the Bayes Least Squares (BLS) estimator of the image $\mathbf{m}$ from the data $\mathbf{d}$, as it minimizes the mean squared estimation error. To regularize as before, we set our inverse prior covariance matrix to our regularization matrix from the standard LSM framework

$$\Lambda^{-1} = \lambda D(\beta) + \varepsilon I \quad (8)$$

so that

$$\hat{\mathbf{m}}_{\text{BLS}}(\mathbf{d}) = \left(A^T \Sigma^{-1} A + \lambda D(\beta) + \varepsilon I\right)^{-1} A^T \Sigma^{-1} \mathbf{d}. \quad (9)$$

Hierarchical Bayesian Formulation

We now describe how we can expand the above formulation to estimate the regularization parameters. In particular, we focus on the problem of estimating the vector of differencing coefficients β from the data $\mathbf{d}$, in addition to the image $\mathbf{m}$. The reason for estimating β in particular is that β captures our prior belief about *where* we think the image should be smooth. In probabilistic terms, β captures the prior conditional dependence structure of the image $\mathbf{m}$, such that $\beta_{ij} = 0$ implies that, prior to observing $\mathbf{d}$, m_i is conditionally independent of m_j when $\{m_k : k \neq i,j\}$ is given. Hence, β defines a Gaussian Markov random field (MRF) on $\mathbf{m}$ prior to observing $\mathbf{d}$; this is depicted in Figure 1 for a simple nine pixel image.

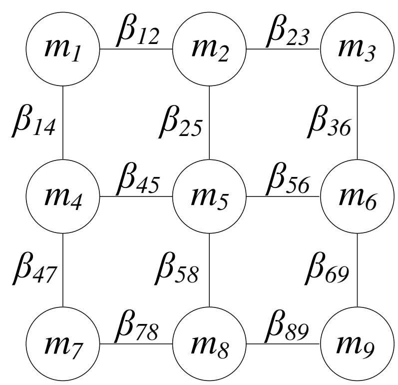

Figure 1: The Markov random field imposed on $\mathbf{m}$ by fixing β prior to observing the data $\mathbf{d}$, for a simple nine pixel image.

In order to estimate β from $\mathbf{d}$, rather than fixing β, we make β a random vector endowed with its own prior $p(\beta)$. This is a reasonable approach since the regularization parameters β give us prior information about our model $\mathbf{m}$, and $\mathbf{m}$ gives us information about our data $\mathbf{d}$, hence we should be able to infer something about β from $\mathbf{d}$. This is depicted in the directed graphical model of Figure 2, which also illustrates the Markov chain structure between β, $\mathbf{m}$, and $\mathbf{d}$. To define our specific probabilistic model for β, $\mathbf{m}$, and $\mathbf{d}$, we endow each β_{ij} with a uniform prior on the set $\{0,1\}$ and let $\{\beta_{ij} : (i,j) \in E\}$ be a set of mutually independent random variables. That is,

$$p(\beta_{ij}) = \frac{1}{2}\delta(\beta_{ij}) + \frac{1}{2}\delta(\beta_{ij} - 1) \quad (10)$$

and

$$p(\beta) = \prod_{(i,j)\in E} p(\beta_{ij}). \quad (11)$$

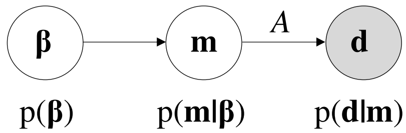

Figure 2: The directed graphical model capturing the Markov chain structure between β, $\mathbf{m}$, and $\mathbf{d}$. The node for $\mathbf{d}$ is shaded to indicate that $\mathbf{d}$ is an observed quantity that the posterior distributions of β and $\mathbf{m}$ are conditioned upon.

Then, conditioning on a particular value for β, we define the probabilistic model for $\mathbf{m}|\beta$ and $\mathbf{d}|\mathbf{m},\beta$ as before, so that

$$\mathbf{m} \,|\, \beta \sim N\left(0, (\lambda D(\beta) + \varepsilon I)^{-1}\right), \tag{12}$$

meaning

$$p(\mathbf{m}|\beta) \propto \frac{\exp\left\{-\frac{1}{2}\mathbf{m}^T(\lambda D(\beta)+\varepsilon I)\mathbf{m}\right\}}{|\lambda D(\beta)+\varepsilon I|^{-1/2}}, \tag{13}$$

and

$$\mathbf{d} \,|\, \mathbf{m}, \beta = \mathbf{d} \,|\, \mathbf{m} \sim N(A\mathbf{m}, \Sigma), \tag{14}$$

meaning

$$\begin{aligned} p(\mathbf{d}|\mathbf{m},\beta) &= p(\mathbf{d}|\mathbf{m}) && (15)\\ &\propto \exp\left\{-\frac{1}{2}(\mathbf{d}-A\mathbf{m})^T\Sigma^{-1}(\mathbf{d}-A\mathbf{m})\right\}. && (16)\end{aligned}$$

We again apply Bayes' rule to obtain the joint posterior distribution for $\mathbf{m}$ *and* β given the data $\mathbf{d}$:

$$\begin{aligned} p(\mathbf{m},\beta|\mathbf{d}) &\propto p(\beta)\,p(\mathbf{m}|\beta)\,p(\mathbf{d}|\mathbf{m},\beta) && (17)\\ &\propto |\lambda D(\beta)+\varepsilon I|^{1/2} \\ &\quad \exp\left\{-\frac{1}{2}\left((\mathbf{d}-A\mathbf{m})^T\Sigma^{-1}(\mathbf{d}-A\mathbf{m}) + \mathbf{m}^T(\lambda D(\beta)+\varepsilon I)\mathbf{m}\right)\right\} && (18)\\ &\quad \prod_{(i,j)\in E}\left(\delta(\beta_{ij})+\delta(\beta_{ij}-1)\right). \end{aligned}$$

We note that Equation 18 is very similar to the posterior distribution in the non-hierarchical Bayesian setting (where β is fixed) with some important differences: firstly, Equation 18 is now a function of both $\mathbf{m}$ *and* β, and secondly, outside the exponential of Equation 18 is the determinant of the regularization matrix. Computing this determinant is expensive, with time complexity $\mathcal{O}(N^3)$ (where N is the number of image parameters), and reflects the additional computational cost of the hierarchical Bayesian approach. The product of delta functions in Equation 18 simply restricts each β_{ij} to being 0 or 1.

Having obtained the posterior distribution $p(\mathbf{m},\beta|\mathbf{d})$, the task of estimating the best image $\mathbf{m}$ remains. We again desire the BLS estimate for $\mathbf{m}$. What is strictly known as the *hierarchical Bayes* solution would be the true BLS estimate of $\mathbf{m}$, the mean of $\mathbf{m}$ on its posterior distribution (marginalizing out β from the joint posterior distribution $p(\mathbf{m},\beta|\mathbf{d})$). Hence, we have for the hierarchical Bayes solution

$$\hat{\mathbf{m}}_{\mathrm{HB,BLS}} = \mathbb{E}[\mathbf{m}|\mathbf{d}] = \int_{\mathscr{M}} \mathbf{m} \int_{\mathscr{B}} p(\mathbf{m},\beta|\mathbf{d})\,\mathrm{d}\beta\,\mathrm{d}\mathbf{m}. \tag{19}$$

One can think of this as searching for the single *best* image $\mathbf{m}$ over *all* choices of regularization parameters β. Unfortunately, the above expectation cannot be given in a closed-form solution and must be computed numerically.

A somewhat different solution to estimating $\mathbf{m}$ is known as the *empirical Bayes* solution, which first looks for the *best* choice for the regularization parameters β_{best} then, using that choice, finds the best image $\hat{\mathbf{m}}_{\mathrm{EB}}$. If one takes the BLS estimate for β then we would have

$$\beta_{\mathrm{best}} = \beta_{\mathrm{BLS}} = \mathbb{E}[\beta|\mathbf{d}] = \int_{\mathscr{B}} \beta \int_{\mathscr{M}} p(\mathbf{m},\beta|\mathbf{d})\,\mathrm{d}\mathbf{m}\,\mathrm{d}\beta \tag{20}$$

where the inner integral can be evaluated analytically, but the outer integral must still be evaluated numerically. We would then obtain the empirical Bayes BLS solution by finding the mean image $\mathbf{m}$ after conditioning on $\beta = \beta_{\mathrm{BLS}}$ so that

$$\begin{aligned} \hat{\mathbf{m}}_{\mathrm{EB,BLS}} &= \mathbb{E}[\mathbf{m}|\mathbf{d},\beta_{\mathrm{BLS}}] && (21)\\ &= \left(A^T\Sigma^{-1}A + \lambda D(\beta_{\mathrm{BLS}}) + \varepsilon I\right)^{-1} A^T\Sigma^{-1}\mathbf{d}. && (22)\end{aligned}$$

Inference Algorithm: Gibbs Sampler

The integrals needed to evaluate the expected values in Equations 19 and 20 (where the integrals over β collapse to summations due to the delta functions) are generally not numerically tractable when the dimensionality of $\mathbf{m}$ (and hence β) are very high. One solution is to approximate the expected values stochastically via sampling. Particularly, if we are able to generate independent samples of a random variable, then we can approximate the expected value of that random variable as the mean of the generated samples. In order to generate samples from $p(\mathbf{m},\beta|\mathbf{d})$, we implement a type of Markov chain Monte Carlo (MCMC) sampler known as the Gibbs sampler. The Gibbs sampler samples a random variable component-by-component from each component's full conditional distribution. Letting $\mathbf{m}_{-i} = \{m_j : j \neq i\}$ and similarly for β_{-ij}, and letting $N = \dim(\mathbf{m})$, the algorithm for the Gibbs sampler to sample from the distribution $p(\mathbf{m},\beta|\mathbf{d})$ is given below:

Initialize $\mathbf{m}^{(0)}, \beta^{(0)}$
for $k = 1, 2, \ldots, K$ **do**
  for $i = 1, 2, \ldots, N$ **do**
    Sample $m_i^{(k)} \sim p\left(m_i \middle| \mathbf{m}_{-i}^{(k-1)}, \beta^{(k-1)}, \mathbf{d}\right)$
  end for
  for all $(i,j) \in E$ **do**
    Sample $\beta_{ij}^{(k)} \sim p\left(\beta_{ij} \middle| \beta_{-ij}^{(k-1)}, \mathbf{m}^{(k)}, \mathbf{d}\right)$
  end for
end for

RESULTS

In order to validate our approach, we ran our inference algorithm on synthetic data sets. We present two test cases: the first arising from an image consisting of three dipping reflectors separated by a weakly reflective fault and the second being a small subsection of the Marmousi model. Synthetic data were created using the same Kirchhoff modeling operator A that is used in the inference algorithms then adding zero-mean white Gaussian noise (having a standard deviation equal to 10% of the maximum amplitude of the data). A homogeneous background velocity model (of 4000 m/s) is used both to create the synthetic data as well as to construct the modeling operator in the inference algorithm. Hence, these test cases are what are known as *inverse crime* tests. The purpose of using the same forward modeling operator to create the synthetic data as is used in the inference is to isolate the *inversion* problem from the *modeling* problem. The variance of the Gaussian noise is also known by the inference algorithm. For both test cases, the data are created from a single surface seismic source (at the center) and 50 equally spaced surface seismic receivers (with spacing of 50 m). The source wavelet is a 20 Hz Ricker wavelet; hence the dominant wavelength (in the 4000 m/s medium) is 200 m. The seismic traces are sampled at 1 ms, and the medium is sampled spatially at 50 m in both the lateral and vertical directions. The entire medium has spatial dimensions of 2500 m by 2500 m, hence $N_x = N_z = 50$ and the number of image parameters is $N = N_x N_z = 2500$.

Figure 3 shows a comparison of the imaging results on the synthetic seismic dataset simulated from the dipping reflectors model and the subsection of the Marmousi model. The results of our methodology are shown in Figures 3(e) and 3(k) (the hierarchical Bayes solutions, where the regularization parameters are integrated out to estimate the image) and 3(f) and 3(l) (the empirical Bayes solutions, where the regularization parameters are estimated from the data then used to estimate the image). A few things are striking about these images. Both the hierarchical and empirical Bayes methods outperform standard LSM imaging, shown in Figures 3(c-d) and 3(i-j). We also notice that the empirical Bayes images seem to do considerably better than the hierarchical Bayes image. This is perhaps due to the fact that hierarchical Bayes typically requires more computation, yet we computed the empirical and hierarchical Bayes images with the *same* number of Gibbs samples. Hence the empirical Bayes result may be more accurate. Comparing the empirical Bayes results to the non-Bayesian results, we notice that we are able to produce a much sharper image of the reflectors; particularly, terminating events tend to smear across the fault in the non-Bayesian images, whereas this is largely remedied in the empirical Bayes images. On the other hand, some of the weaker reflectors in the Marmousi model tend not to appear strongly in the Bayesian imaging results; this is likely due to weak variations in the data being confused for noise.

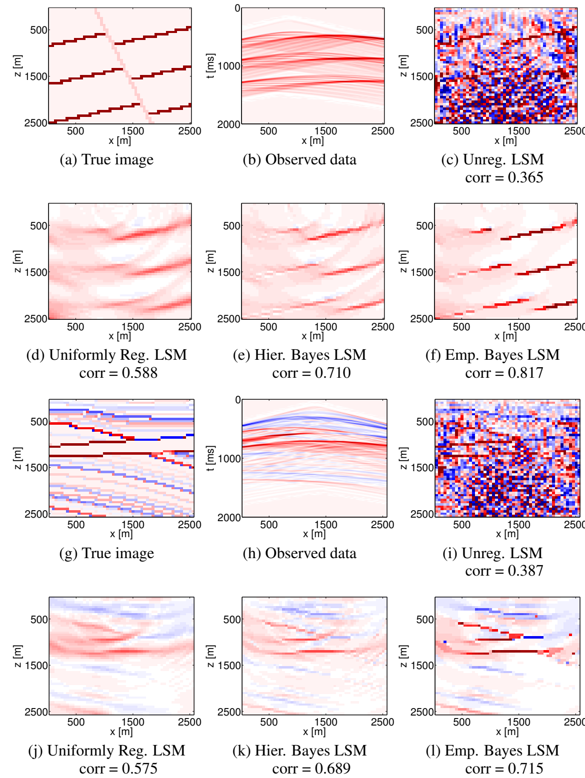
(a) True image; (b) Observed data; (c) Unreg. LSM corr = 0.365; (d) Uniformly Reg. LSM corr = 0.588; (e) Hier. Bayes LSM corr = 0.710; (f) Emp. Bayes LSM corr = 0.817; (g) True image; (h) Observed data; (i) Unreg. LSM corr = 0.387; (j) Uniformly Reg. LSM corr = 0.575; (k) Hier. Bayes LSM corr = 0.689; (l) Emp. Bayes LSM corr = 0.715

Figure 3: (a,g) The true image and (b,h) resulting (single-source) synthetic seismic data with additive noise for the dipping reflectors model and Marmousi subsection model, respectively. Imaging results from (c,i) unregularized LSM, (d,j) regularized LSM using a uniform regularization scheme (i.e. each $\beta_{ij} = 1$), (e,k) hierarchical Bayesian imaging ($\mathbb{E}[\mathbf{m}|\mathbf{d}]$), (f,l) empirical Bayesian imaging ($\mathbb{E}[\mathbf{m}|\beta_{\text{BLS}}, \mathbf{d}]$). Below each image is the correlation of the image with the true image.

CONCLUSIONS AND FUTURE WORK

Our study shows that the Bayesian framework provides a flexible methodology for estimating both the image and regularization parameters in a least-squares migration setting. By estimating where to regularize, we are able to remove the effects of noise while, by and large, preserving sharpness at the reflectors in the image. We note that our current formulation only detects *where* rather than *how strongly* to regularize, and as such tends to work better when different reflectors have comparable strengths. We saw that variations in the data due to weak reflectors may be confused for noise and hence weaker reflectors may not appear as sharply. A natural path for future work would be to generalize this framework to also detect the regularization strength. To allow regularization strength to vary spatially, we could instead allow β_{ij} to be any non-negative number rather than restricted to 0 or 1. Another issue which remains to be addressed is the computational complexity of our algorithm; we would like to be able to make this methodology scalable to 3-D or large 2-D data sets. Hence another avenue for future work is to improve the efficiency of the algorithm.

ACKNOWLEDGEMENTS

This work was supported by Shell International E&P, Inc. and the ERL Founding Member Consortium. The authors thank Vanessa Goh, Ken Matson, and Henning Kuehl of Shell for insightful discussions and feedback.

REFERENCES

Bleistein, N., 1984, Mathematical methods for wave phenomena: Academic Press. Computer science and applied mathematics.

Bodin, T., M. Sambridge, N. Rawlinson, and P. Arroucau, 2012, Transdimensional tomography with unknown data noise: Geophysical Journal International, **189**, 1536–1556.

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.

Claerbout, J., 1992, Earth soundings analysis: Processing versus inversion: Blackwell Scientific Publ. Stanford Exploration project.

Clapp, M., 2005, Imaging under salt: illumination compensation by regularized inversion: PhD thesis, Stanford University.

Duquet, B., K. J. Marfurt, and J. A. Dellinger, 2000, Kirchhoff modeling, inversion for reflectivity, and subsurface illumination: Geophysics, **65**, 1195–1209.

Green, P. J., 1995, Reversible jump markov chain monte carlo computation and bayesian model determination: Biometrika, **82**, 711–732.

Hanitzsch, C., J. Schleicher, and P. Hubral, 1994, True-amplitude migration of 2d synthetic data1: Geophysical Prospecting, **42**, 445–462.

Kühl, H., and M. D. Sacchi, 2003, Least-squares wave-equation migration for avp/ava inversion: Geophysics, **68**, 262–273.

Lambare, G., J. Virieux, R. Madariaga, and S. Jin, 1992, Iterative asymptotic inversion in the acoustic approximation: Geophysics, **57**, 1138–1154.

LeBras, R., and R. W. Clayton, 1988, An iterative inversion of back-scattered acoustic waves: Geophysics, **53**, 501–508.

Malinverno, A., 2000, A bayesian criterion for simplicity in inverse problem parametrization: Geophysical Journal International, **140**, 267–285.

———, 2002, Parsimonious bayesian markov chain monte carlo inversion in a nonlinear geophysical problem: Geophysical Journal International, **151**, 675–688.

Malinverno, A., and V. A. Briggs, 2004, Expanded uncertainty quantification in inverse problems: Hierarchical bayes and empirical bayes: Geophysics, **69**, 1005–1016.

Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208–221.