Bayesian wavefield separation by transform-domain sparsity promotion

Deli Wang*, Rayan Saab†, Özgür Yılmaz‡ and Felix J. Herrmann§

(March 13, 2008)

Running head: **Curvelet-based primary-multiple separation**

ABSTRACT

Successful removal of coherent noise sources greatly determines the quality of seismic imaging. Major advances were made in this direction, e.g., Surface-Related Multiple Elimination (SRME) and interferometric ground-roll removal. Still, moderate phase, timing, amplitude errors and clutter in the predicted signal components can be detrimental. Adopting a Bayesian approach along with the assumption of approximate curvelet-domain independence of the to-be-separated signal components, we construct an iterative algorithm that takes the predictions produced by for example SRME as input and separates these components in a robust fashion. In addition, the proposed algorithm controls the energy mismatch between the separated and predicted components. Such a control, which was lacking in earlier curvelet-domain formulations, produces improved results for primary-multiple separation on both synthetic and real data.

*College of Geoexploration Science and Technology, Jilin University, 6 Ximinzhu street, Changchun, 130026, China. Visiting the University of British Columbia.

†Department of Electrical and Computer Engineering, University of British Columbia

‡Department of Mathematics, University of British Columbia

§Seismic Laboratory for Imaging and Modeling, Department of Earth and Ocean Sciences, University of British Columbia, 6339 Stores Road, Vancouver, V6T 1Z4, BC, Canada

INTRODUCTION

Successful separation of coherent signal components constitutes an important step in the seismic processing flow. Coherent noise separation, such as primary-multiple separation as part of Surface-Related Multiple Elimination (SRME) (Verschuur et al., 1992; Fokkema and van den Berg, 1993; Berkhout and Verschuur, 1997; Weglein et al., 1997) or interferometric ground-roll removal (Dong and Schuster, 2006), typically involves two stages: a noise-prediction stage and a primary-reflection noise separation stage. During the second stage, measures need to be taken to compensate for moderate kinematic and dynamic errors in the predictions of the noise components (see e.g. Verschuur et al., 1992; Ikelle et al., 1997). For instance, 3-D complexity in the Earth produces these types of errors during SRME (Verschuur et al., 1992; Ikelle et al., 1997; Dragoset and Jeričević, 1998; Ross et al., 1999; Verschuur, 2006), while multiple (van Borselen et al., 2004; Lin et al., 2004; Moore and Dragoset, 2004; van Dedem and Berkhout, 2005) and interferometric ground-roll predictions (Dong and Schuster, 2006) become unreliable for incomplete data acquisitions. As a result, separation methods based on conventional (windowed) convolutional matched filtering may not compensate adequately for these errors, rendering noise removal ineffective.

Our main concern in this letter is the construction of an iterative separation algorithm that is stable with respect to moderate errors in the noise predictions. The traditional "separation by subtraction after matching" approach yields predictions of signal components with residual amplitude, timing, dip, phase, and wavelet errors. The separation obtained via such methods can be improved via an additional curvelet-domain nonlinear matching procedure that can handle significant amplitude errors provided they vary smoothly as a function of the location and dips of the noise predictions (Herrmann et al., 2007b). Assuming that the

resulting predictions (possibly after this additional matched filtering stage) have moderate residual errors, the algorithm proposed in this paper complements such ℓ_2-matched-filtering techniques where noise in the total data and errors in the noise predictions lead to residual noise energy, high-frequency clutter, and deterioration of primary energy (Chen et al., 2004; Abma et al., 1005; Herrmann et al., 2007a). In particular, our algorithm takes the predictions obtained using the traditional methods as input, and produces improved estimates of the primaries and multiples. (For simplicity, we will refer to such predictions as the "SRME-predictions" although our algorithm is generic and the predictions could be obtained using any method appropriate for the type of wavefield-separation problem.) This is achieved by exploiting the wavefront detection and hence separation capability of curvelets (Herrmann et al., 2007a, 2008) while offering additional control over the energy mismatch between the SRME-predicted noise components (the input) and the estimated signal components (the output). Previous separation algorithms based on curvelet-domain sparsity lack this type of control (Herrmann et al., 2007a, 2008). This letter builds explicitly on more general results on transform-based signal separation that will be reported elsewhere.

THEORY

The proposed method involves the separation of two (or more since our method extends to more than two signal components) coherent signal components, given a prediction for one of the signal components. In SRME, this problem corresponds to separating primaries and multiples, given a prediction for the multiples with moderate errors. To solve this problem, we present a Bayesian formulation that allows for errors in the total data as well as in the predicted noise component. Without loss of generality, the ensuing discussion is limited to the problem of primary-multiple separation, where the total data and predicted multiples

serve as input and during which estimates for the primaries and multiples are calculated. The predicted multiples are assumed to be given by the matched filter of standard SRME, possibly supplemented by a curvelet-domain matched filter recently proposed by Herrmann et al. (2007b).

The forward model

The primary-multiple separation problem is cast into a probabilistic framework where the recorded total data vector $\mathbf{b}$,

$$\mathbf{b} = \mathbf{s}_1 + \mathbf{s}_2 + \mathbf{n}, \tag{1}$$

is assumed to consist of a superposition of primaries, $\mathbf{s}_1$, multiples, $\mathbf{s}_2$, and white Gaussian noise $\mathbf{n}$, each component of which is $N(0, \sigma^2)$—i.e., zero-mean Gaussian with standard deviation σ. We denote by $\mathbf{b}_2$ the vector that consists of the SRME-predicted multiples. As the SRME predictions contain errors, we assume

$$\mathbf{b}_2 = \mathbf{s}_2 + \mathbf{n}_2, \tag{2}$$

where $\mathbf{n}_2$, the error in the predicted multiples is also assumed white Gaussian; each component of $\mathbf{n}_2$ is $N(0, \sigma_2^2)$. Moreover we assume that $\mathbf{n}$ and $\mathbf{n}_2$ are independent.

Following earlier work (Herrmann et al., 2007a), we write the two unknown signal components as a superposition of curvelets—i.e., $\mathbf{s_i} = \mathbf{A}\mathbf{x}_i,\ i = 1, 2$, where $\mathbf{A}$ is the curvelet-synthesis matrix (Candes et al., 2006), and obtain the system of equations

$$\begin{aligned} \mathbf{b}_1 &= \mathbf{A}\mathbf{x}_1 + \mathbf{n}_1, \\ \mathbf{b}_2 &= \mathbf{A}\mathbf{x}_2 + \mathbf{n}_2. \end{aligned} \tag{3}$$

Here, $\mathbf{n}_1 = \mathbf{n} - \mathbf{n}_2$. Moreover, the unknown curvelet coefficients for the primaries, $\mathbf{x}_1$, and multiples, $\mathbf{x}_2$, are related to the SRME-predicted primaries, $\mathbf{b}_1 = \mathbf{b} - \mathbf{b}_2$, and SRME-

predicted multiples, $\mathbf{b}_2$. With this formulation, we are in a position to exploit the sparsity of the curvelet coefficient vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ for separating the two signal components.

Remark: Note that our assumption that $\mathbf{n}$ in Equation 1 is white Gaussian noise is consistent with the assumptions underlying the matched filter used in the SRME-multiple prediction, see (Verschuur et al., 1992). Furthermore, our formulation can be extended to the case where both noise contributions $\mathbf{n}$ and $\mathbf{n}_2$ are colored Gaussian as long as the corresponding covariance matrices are known. In the absence of an accurate probabilistic model for the prior distribution of the error in the predicted multiples, we make the simplifying assumption that this error is Gaussian. As we will show in the next section, this assumption leads to an optimization problem with a ℓ_2-norm penalty for the error. In this way, we merely assign a cost function that penalizes the ℓ_2 misfit between the predicted multiples and the estimates produced by our algorithm. In addition, the fidelity of the predicted multiples can be incorporated by choosing the parameters in this optimization procedure appropriately.

Bayesian signal separation

Given the above formulation of the forward model, (cf. Equation 3), we now derive conditional probabilities for the unknown curvelet coefficient vectors, which we initially assume to be given by realizations of two independent random processes with weighted Laplacian-type probability density functions. Such a choice serves as a sparsity-promoting prior (see e.g. Zibulevsky and Pearlmutter (2001); Li et al. (2004), and Taylor et al. (1979); Oldenburg et al. (1981); Ulrych and Walker (1982) in the geophysical literature), and is consistent with the high compression rates that curvelets attain on seismic data (Candes et al., 2006; Hen-

nenfent and Herrmann, 2006; Herrmann et al., 2007a). Given the SRME predictions—i.e., $\mathbf{b}_1$ and $\mathbf{b}_2$, our goal is to estimate the curvelet coefficients for the two signal components—i.e., $\mathbf{x}_1$ and $\mathbf{x}_2$. Probabilistically, this means that our objective is to find the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ that maximize the conditional probability $P(\mathbf{x}_1, \mathbf{x}_2|\mathbf{b}_1, \mathbf{b}_2)$. In other words, using Bayes' rule we need to maximize

$$P(\mathbf{x}_1, \mathbf{x}_2|\mathbf{b}_1, \mathbf{b}_2) = \frac{P(\mathbf{x}_1, \mathbf{x}_2)P(\mathbf{b}_1|\mathbf{x}_1, \mathbf{x}_2)P(\mathbf{b}_2|\mathbf{b}_1, \mathbf{x}_1, \mathbf{x}_2)}{P(\mathbf{b}_1, \mathbf{b}_2)} = \frac{P(\mathbf{x}_1, \mathbf{x}_2)P(\mathbf{n})P(\mathbf{n}_2)}{P(\mathbf{b}_1, \mathbf{b}_2)}. \quad (4)$$

Since both $\mathbf{b}_1$ and $\mathbf{b}_2$ are known, we try to find $\widetilde{\mathbf{x}}_1$ and $\widetilde{\mathbf{x}}_2$, the curvelet coefficients for the primaries and multiples, that maximize the posterior probability in Equation 4 under the assumptions: (i) $\mathbf{n}$ and $\mathbf{n}_2$ are independent white Gaussian noise vectors with possibly different variances as described above, and (ii) $\mathbf{x}_1$ and $\mathbf{x}_2$ have weighted Laplacian prior distributions. More precisely, $\widetilde{\mathbf{x}}_1$ and $\widetilde{\mathbf{x}}_2$ solve the optimization problem

$$\begin{aligned}
&\max_{\mathbf{x_1},\mathbf{x}_2} P(\mathbf{x}_1, \mathbf{x}_2|\mathbf{b}_1, \mathbf{b}_2) \\
&\quad = \max_{\mathbf{x_1},\mathbf{x}_2} \; P(\mathbf{x}_1, \mathbf{x}_2)P(\mathbf{n})P(\mathbf{n}_2) \\
&\quad = \max_{\mathbf{x_1},\mathbf{x}_2} \; \exp\left(-\alpha_1\|\mathbf{x}_1\|_{1,\mathbf{w}_1} - \alpha_2\|\mathbf{x}_2\|_{1,\mathbf{w}_2} - \frac{\|\mathbf{A}\mathbf{x}_2 - \mathbf{b}_2\|_2^2}{\sigma_2^2} - \frac{\|\mathbf{A}(\mathbf{x}_1 + \mathbf{x}_2) - (\mathbf{b}_1 + \mathbf{b}_2)\|_2^2}{\sigma^2}\right) \\
&\quad = \max_{\mathbf{x_1},\mathbf{x}_2} -\left(\alpha_1\|\mathbf{x}_1\|_{1,\mathbf{w}_1} + \alpha_2\|\mathbf{x}_2\|_{1,\mathbf{w}_2} + \frac{\|\mathbf{A}\mathbf{x}_2 - \mathbf{b}_2\|_2^2}{\sigma_2^2} + \frac{\|\mathbf{A}(\mathbf{x}_1 + \mathbf{x}_2) - (\mathbf{b}_1 + \mathbf{b}_2)\|_2^2}{\sigma^2}\right), \quad (5)
\end{aligned}$$

yielding estimates $\widetilde{\mathbf{s}}_1 = \mathbf{A}\widetilde{\mathbf{x}}_1$ for the primaries, and $\widetilde{\mathbf{s}}_2 = \mathbf{A}\widetilde{\mathbf{x}}_2$ for the multiples. With appropriate rescaling, Equation 5 reduces to $\max_{\mathbf{x_1},\mathbf{x}_2} P(\mathbf{x}_1, \mathbf{x}_2|\mathbf{b}_1, \mathbf{b}_2) = \min_{\mathbf{x_1},\mathbf{x}_2} f(\mathbf{x}_1, \mathbf{x}_2)$ with

$$f(\mathbf{x}_1, \mathbf{x}_2) = \lambda_1\|\mathbf{x}_1\|_{1,\mathbf{w}_1} + \lambda_2\|\mathbf{x}_2\|_{1,\mathbf{w}_2} + \|\mathbf{A}\mathbf{x}_2 - \mathbf{b}_2\|_2^2 + \eta\|\mathbf{A}(\mathbf{x}_1 + \mathbf{x}_2) - \mathbf{b}\|_2^2. \quad (6)$$

Here $\|\mathbf{x}_i\|_{1,\mathbf{w}_i} = \sum_{\mu\in\mathcal{M}} |w_{i,\mu}x_{i,\mu}|$ is the weighted ℓ^1-norm of the curvelet coefficients $\mathbf{x}_i$ (and $\mathcal{M}$ is the index set for the curvelet coefficients). Heuristically, the parameters λ_1 and λ_2 control the tradeoff between the sparsity of the coefficient vectors and the misfit with respect to the SRME-predictions (the total data and the multiple predictions). On the other hand, η controls the tradeoff between the confidence in the total data and in the SRME-predicted multiples.

The separation algorithm

For appropriately chosen $\lambda_1, \lambda_2, \eta$, and reasonably accurate SRME-predicted multiples, the minimization of the objective function $f(\mathbf{x}_1, \mathbf{x}_2)$ leads to a separation of the primaries and multiples. To minimize $f(\mathbf{x}_1, \mathbf{x}_2)$ in Equation 6, we devise an iterative thresholding algorithm in the spirit of the work by Daubechies et al. (2003). Starting from arbitrary initial estimates $\mathbf{x}_1^0$ and $\mathbf{x}_2^0$ of $\mathbf{x}_1$ and $\mathbf{x}_2$, the n^{th} iteration of the algorithm proceeds as follows

$$\begin{aligned} \mathbf{x}_1^{n+1} &= \mathbf{T}_{\frac{\lambda_1 \mathbf{w_1}}{2\eta}} \left[\mathbf{A}^T\mathbf{b}_2 - \mathbf{A}^T\mathbf{A}\mathbf{x}_2^n + \mathbf{A}^T\mathbf{b}_1 - \mathbf{A}^T\mathbf{A}\mathbf{x}_1^n + \mathbf{x}_1^n\right] \\ \mathbf{x}_2^{n+1} &= \mathbf{T}_{\frac{\lambda_2 \mathbf{w_2}}{2(1+\eta)}} \left[\mathbf{A}^T\mathbf{b}_2 - \mathbf{A}^T\mathbf{A}\mathbf{x}_2^n + \mathbf{x}_2^n + \tfrac{\eta}{\eta+1}\left(\mathbf{A}^T\mathbf{b}_1 - \mathbf{A}^T\mathbf{A}\mathbf{x}_1^n\right)\right], \end{aligned} \tag{7}$$

where $\mathbf{T_u} : \mathbb{R}^{|\mathcal{M}|} \mapsto \mathbb{R}^{|\mathcal{M}|}$ is the elementwise soft-thresholding operator—i.e., for each $\mu \in \mathcal{M}$, $T_{u_\mu}(v_\mu) := \text{sgn}(v_\mu) \cdot \max(0, |v_\mu| - |u_\mu|)$. The proposed algorithm provably converges to the minimizer of $f(\mathbf{x}_1, \mathbf{x}_2)$, provided all weights—i.e., all components of the vectors $\mathbf{w}_1$ and $\mathbf{w}_2$, are strictly positive (Daubechies et al., 2003).

Empirical choice of the weights

In practice, the appropriate choice of the weights $\mathbf{w}_1$ and $\mathbf{w}_2$ proves important. Motivated by empirical findings (Herrmann et al., 2007a), we choose the weights using the SRME predictions for the two signal components: we set $\mathbf{w}_1 = \max\{|\mathbf{A}^T\mathbf{b}_2|, \epsilon\}$ and $\mathbf{w}_2 = \max\{|\mathbf{A}^T\mathbf{b}_1|, \epsilon\}$ with the operations taken elementwise. Here, $\mathbf{A}^T$ is the forward curvelet transform (the discrete curvelet transform based on wrapping is a tight frame, so the transpose of this transform is its inverse) and ϵ a noise dependent constant. This choice of weights guarantees that the weights are strictly positive, thus the algorithm converges. Furthermore, such a choice makes it less likely that the optimization algorithm produces large curvelet

coefficients for the primaries at entries in the curvelet vector that exhibit large coefficients for the predicted multiples, and vice versa.

Remark: Within the Bayesian context, this particular choice breaks the assumption that the curvelet coefficients of the primaries and multiples are independent. However, our estimator can still be interpreted as Bayesian with a posterior probability of the form given in Equation 5 with the weights defined as mentioned above.

DISCUSSION

As shown in our derivation, the Bayesian formulation in Equation 5 leads to an optimization problem—i.e., the minimization of $f(\mathbf{x}_1, \mathbf{x}_2)$, which involves a combined minimization of the weighted ℓ^1-norms of the coefficient vectors, the ℓ^2-misfit between the SRME-predicted and estimated multiples, and the ℓ^2-misfit between the sum of the estimated primaries and multiples and the observed total data. From this interpretation, it is clear that our Bayesian formulation is an extension of earlier work (Herrmann et al., 2007a) since our objective function includes an additional term. This new term acts as a safeguard by making sure that the estimated multiples remain sufficiently close to the SRME-predicted multiples.The lower our confidence is in the SRME-predicted multiples, the more emphasis we place on the total data misfit. The case $\eta \rightarrow \infty$ is analogous to an absolute lack of confidence on the SRME-predicted multiples, and thus includes a misfit concerning the total data only. This limiting unrealistic assumption underlaid our earlier formulation (Herrmann et al., 2007a). Away from this limit, Equation 6 leads to solutions that are not only sparse, but also produce estimated curvelet coefficients for the multiples that are required to fit the SRME-predicted multiples. The relative degrees of sparsity for the two signal components are controlled by λ_1 and λ_2.

The performance of the presented separation algorithm depends on: (i) the sparsity of the coherent signal components in the transform domain: the sparser the two signal components, the smaller the chance that the supports of the two curvelet coefficient vectors overlap; (ii) the validity of the independence assumption, which is empirically established in Herrmann et al. (2007a); (iii) the accuracy of the SRME-predictions. Even though it was shown that curvelet-domain separation is relatively insensitive to errors in the SRME-predictions, significant amplitude errors (significant timing errors are assumed absent) lead to a deterioration of the separation. However, for smoothly varying amplitude errors, a remedy for this situation has recently been proposed that is based on introducing a secondary curvelet-domain matched filter (Herrmann et al., 2007b).

EXAMPLES

Synthetic data example: In Figure 1, we make comparisons between primaries derived from the total data in Figure 1(a) using standard SRME (Figure 1(c)), 3-D curvelets with single thresholding (Figure 1(d)), and our Bayesian method with (Figure 1(e)) and without control on the estimated multiples (Figure 1(f)). For reference, we included "multiple-free" data in Figure 1(b). This data was modeled with an absorbing boundary condition, removing the surface-related multiples. The SRME-predicted primaries plotted in Figure 1(c) were obtained with a multiple-window matched-filtering procedure. Comparing these results with the curvelet-domain separations shows significant improvements. This behavior is confirmed by signal-to-noise ratios (SNRs) that measure the performance of our algorithm for various parameter settings. To accommodate lack of correct scaling for the different

signal components, we define the SNR as

$$\mathrm{SNR} = 20 \log_{10} \frac{\| \frac{\mathbf{s}_1}{\|\mathbf{s}_1\|_2} \|_2}{\| \frac{\tilde{\mathbf{s}}_1}{\|\tilde{\mathbf{s}}_1\|_2} - \frac{\mathbf{s}_1}{\|\mathbf{s}_1\|_2} \|_2}, \tag{8}$$

where the energies of the 'multiple-free' data, $\mathbf{s}_1$, and the estimated primaries, $\widetilde{\mathbf{s}}_1$, are both normalized to one. First, Bayesian separation with optimal parameters yields an SNR of 12.13 dB compared to 10.27 dB for single thresholding. These SNRs are computed with respect to the "multiple-free" data. Second, the result for $\lambda_1^* = 0.7$, $\lambda_2^* = 2.0$ and $\eta^* = 0.5$ in Figure 1(e) shows less remnant multiple energy compared to the standard SRME result with an SNR of 9.82 dB. These parameter choices were found empirically and the smaller value for λ_1^* is consistent with the predicted multiples and hence the weights $\mathbf{w}_1$ being less sparse than the predicted primaries. For this example, the solution converged in only five iterations of Equation 7, which is quite remarkable given the extreme problem size ($> 2^{31}$ unknowns). Table 1 summarizes the performance in SNRs compared to the ground truth shown in Figure 1(b). These numbers lead us to conclude that our separation scheme is relatively robust against changes in the control parameters (cf. the value for $\{\lambda_1^*, \lambda_2^*, \eta^*\}$ compared to other choices) and that control over the estimated multiples ($\{\lambda_1^*, \lambda_2^*, \eta^*\}$ versus reduced control for $100 \cdot \{\lambda_1^*, \lambda_2^*, \eta^*\}$) leads to better results. However, our method leads to minor loss of primary energy especially in regions of high curvature. By changing the parameters, this loss can be reduced at the expense of more residual multiple energy. For a more detailed study of this example, refer to the ancillary material which includes difference plots. A more fundamental solution to this problem will be investigated in future work.

Field data example: Figure 2(a) contains the common-offset section (at offset 200 m) that we selected from a North Sea field dataset. SRME-predicted multiples and primaries

SNR (dB)	$\{\lambda_1^*,\, \lambda_2^*\}$	$\{2 \cdot \lambda_1^*,\, \lambda_2^*\}$	$\{\lambda_1^*,\, 2 \cdot \lambda_2^*\}$	$100 \cdot \{\lambda_1^*,\, \lambda_2^*\}$
η^*	12.13	11.21	11.46	-
$\frac{1}{2} \cdot \eta^*$	11.36	9.43	11.46	-
$2 \cdot \eta^*$	11.44	12.13	9.92	-
$100 \cdot \eta^*$	-	-	-	10.65

Table 1: Sensitivity analysis for the performance of Bayesian separation for different sparsity ratios (λ_1/λ_2), and different levels of fidelity in the predicted multiples η^{-1}. Here, $(\lambda_1^*,\, \lambda_2^*,\, \eta^*) = (0.7,\, 2.0,\, 0.5)$. The SNRs are computed via Equation 8 and are relatively robust against changes in parameters. Parameter combinations for which the ratios of the sparsity over level of fidelity are not preserved were omitted since these lead to extremely low SNRs. Note that the inclusion of control on the estimated multiples, yielding an SNR of 12.13 dB, adds 1.48 dB to the estimated primaries without this control, yielding an SNR of 10.65 dB.

are plotted in Figure 2(b) and 2(c). Comparison of the SRME-predicted primaries with the section obtained by our Bayesian separation method using the 3-D curvelet transform shows a clear improvement in the removal of shallow multiples and improved continuity for late primary arrivals. These results were obtained for $\lambda_1 = 1.0$, $\lambda_2 = 2.0$ and $\eta = 1.0$. These parameter choices are consistent with less-than-ideal real data for which we can not expect significant changes in sparsity between primaries and multiples and for which we can not completely trust the predicted multiples. In this example, the solution is attained after only ten iterations. Again, refer to the ancillary material for a more detailed study of this

example.

CONCLUSIONS

In this letter, we introduce a robust algorithm for the separation of coherent signal components that are sparse in the same transform domain. The separation problem is formulated in terms of Bayesian statistics where curvelet-domain sparsity and approximate independence of the to-be-separated signal components both serve as priors. Given initial predictions of signal components that contain moderate errors, the proposed algorithm outputs improved estimates of these components. Convergence of our separation algorithm is assured by defining the weighted one-norms of the signal components in terms of the initial signal predictions that serve as input. The fast convergence and the quality of the separation results both follow from the ability of curvelets to sparsely represent each signal component. This observation opens the tantalizing perspective of a generic algorithm where coherent signal components are successfully separated, given signal predictions with moderate errors. Our excellent results on primary-multiple separation seem to underline this perspective.

ACKNOWLEDGMENTS

The authors would like to thank Eric Verschuur for his input in the primary-multiple separation. We also would like to thank the authors of CurveLab for making their codes available. The examples presented were prepared with Madagascar supplemented by SLIMPy operator overloading, developed by Sean Ross-Ross and Henryk Modzelewski. Norsk Hydro is thanked for making the field dataset available. R.S. was partly supported by the Province

of British Columbia through the Ministry of Advanced Education with a Pacific Century Graduate Scholarship. In addition, this work was in part financially supported by the NSERC Discovery Grant (22R81254) of F.J.H. and CRD Grant DNOISE (334810-05) of O.Y. and F.J.H., and was carried out as part of the SINBAD project with support, secured through ITF, from BG Group, BP, Chevron, ExxonMobil and Shell.

REFERENCES

Abma, R., N. K. amnd K. H. Matson, S. Michell, S. A. Shaw, and B. McLain, 1005, Comparisons of adaptive subtraction methods for multiple attenuation: The Leading Edge, **24**, 277–280.

Berkhout, A. J. and D. J. Verschuur, 1997, Estimation of multiple scattering by iterative inversion, part I: theoretical considerations: Geophysics, **62**, 1586–1595.

Candes, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006, Fast discrete curvelet transforms: SIAM Multiscale Model. Simul., **5**, 861–899.

Chen, J., E. Baysal, and O. Yilmaz, 2004, Weighted subtraction for diffracted multiple attenuation: 74th Ann. Internat. Mtg., SEG, Expanded Abstracts, 1329–1332, Soc. Expl. Geophys., Expanded abstracts.

Daubechies, I., M. Defrise, and C. De Mol, 2003, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint.

Dong, S., R. H. and G. Schuster, 2006, Interferometric prediction and least-squares subtraction of surface waves: 76th Annual International Meeting, SEG, Expanded Abstracts, 2783–2786.

Dragoset, W. H. and Z. Jeričević, 1998, Some remarks on multiple attenuation: Geophysics, **63**, 772–789.

Fokkema, J. T. and P. M. van den Berg, 1993, Seismic applications of acoustic reciprocity: Elsevier.

Hennenfent, G. and F. J. Herrmann, 2006, Seismic denoising with non-uniformly sampled curvelets: IEEE Comp. in Sci. and Eng., **8**, 16–25.

Herrmann, F., D. Wang, G. Hennenfent, and P. Moghaddam, 2008, Curvelet-based seismic data processing: a multiscale and nonlinear approach: Geophysics, **73**, A1–A5.

(doi:10.1190/1.2799517).

Herrmann, F. J., U. Boeniger, and D. J. Verschuur, 2007a, Nonlinear primary-multiple separation with directional curvelet frames: Geoph. J. Int., **170**, 781–799. (doi: 10.1111/j.1365-246X.2007.03360.x).

Herrmann, F. J., D. Wang, and D. J. Verschuur, 2007b, Adaptive curvelet-domain primary-multiple separation. (Accepted for publication in Geophysics GEO-2007-0234).

Ikelle, L., G. Roberts, and A. Weglein, 1997, Source signature estimation based on the removal of first-order multiples: Geophysics, **62**, 1904–1920.

Li, Y., A. Cichocki, and S. Amari, 2004, Analysis of sparse representation and blind source separation: Neural Compututation, **16**, 1193–1234.

Lin, D., J. Young, Y. Huang, and M. Hartmann, 2004, 3-D srme application in the gulf of mexico: 74th Ann. Internat. Mtg., SEG, Expanded Abstracts, 1257–1260, Soc. Expl. Geophys., Expanded abstracts.

Moore, I. and W. Dragoset, 2004, 3-D surface-related multiple prediction (smp): 74th Ann. Internat. Mtg., SEG, Expanded Abstracts, 1249–1252, Soc. Expl. Geophys., Expanded abstracts.

Oldenburg, D. W., S. Levy, and K. P. Whittall, 1981, Wavelet estimation and deconvolution: Geophysics, **46**, 1528–1542.

Ross, W. S., Y. Yu, and F. A. Gasparotto, 1999, Traveltime prediction and suppression of 3-D multiples: Geophysics, **64**, 261–277.

Taylor, H. L., S. Banks, and J. McCoy, 1979, Deconvolution with the ℓ_1 norm: Geophysics, **44**, 39.

Ulrych, T. J. and C. Walker, 1982, Analytic minimum entropy deconvolution: Geophysics, **47**, 1295–1302.

van Borselen, R., R. Schonewille, and R. Hegge, 2004, 3-D srme: Acquisition and processing solutions: 74th Ann. Internat. Mtg., SEG, Expanded Abstracts, 1241–1244, Soc. Expl. Geophys., Expanded abstracts.

van Dedem, E. J. and A. J. Berkhout, 2005, 3-D surface-related multiple prediction: A sparse inversion approach: Geophysics, **70**, V31–V43.

Verschuur, D. J., 2006, Seismic multiple removal techniques: past, present and future, EAGE publications b.v. ed.

Verschuur, D. J., A. J. Berkhout, and C. P. A. Wapenaar, 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, 1166–1177.

Weglein, A. B., F. A. Carvalho, and P. M. Stolt, 1997, An iverse scattering series method for attenuating multiples in seismic reflection data: Geophysics, **62**, 1975–1989.

Zibulevsky, M. and B. A. Pearlmutter, 2001, Blind source separation by sparse decomposition in a signal dictionary: Neural Computation, **13**, 863–882.

LIST OF FIGURES

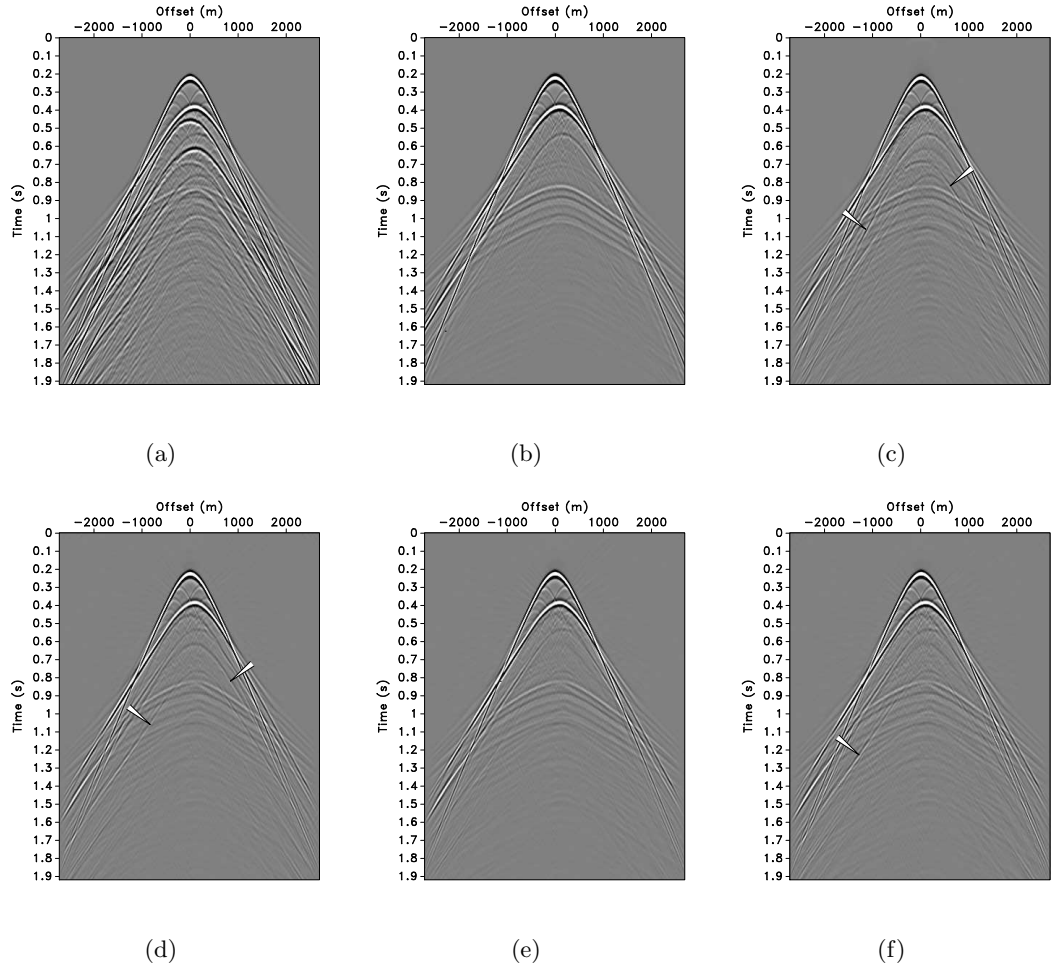

Figure 1: Primary-multiple separation on a synthetic data volume. **(a)** The total data, **b**. **(b)** Reference surface-related multiple-free data modeled with an absorbing boundary condition. **(c)** SRME-predicted primaries, $\mathbf{b}_1$. **(d)** Estimate for the primaries, using 3-D curvelets and single thresholding. **(e)** The same but with Bayesian thresholding for $\lambda_1^* = 0.7$, $\lambda_2^* = 2.0$ and $\eta^* = 0.5$. **(f)** The same as **(e)** but now for $\{\lambda_1, \lambda_2, \eta\} = 100 \cdot \{\lambda_1^*, \lambda_2^*, \eta^*\}$. Notice the improvement in the estimated primaries by controlling the estimated multiples. By multiplying the η and the other control parameters by a large factor, we diminished the control over the multiple prediction while keeping the sparsity penalties the same. Less control clearly adversely affects the estimated primaries, which is confirmed by the SNRs computed with respect to **(b)**, i.e, 12.13 dB versus 10.65 dB.

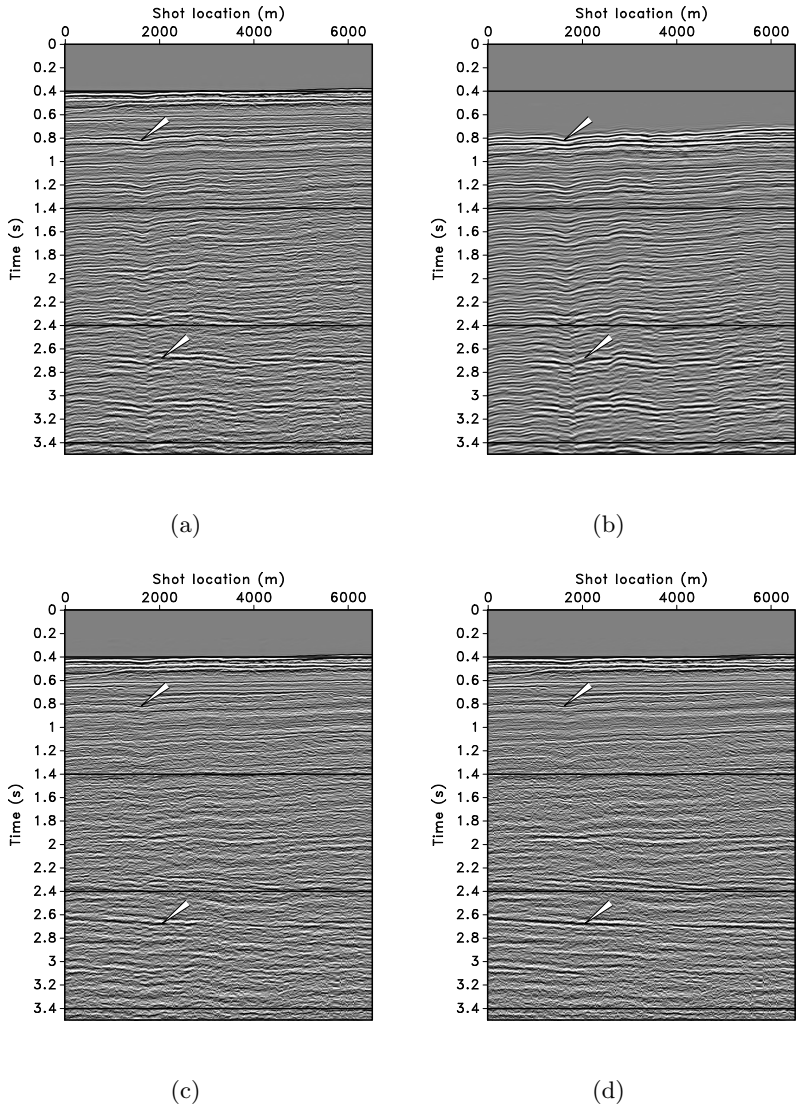

Figure 2: Field data example of curvelet-domain primary-multiple separation. **(a)** Near-offset (200 m) section for the total data plotted with automatic-gain control. **(b)** Estimate for the multiples, yielded by optimized multi-window SRME. **(c)** Corresponding estimate for the primaries using SRME. **(d)** Estimate for the primaries computed by Bayesian iterative thresholding with $\lambda_1 = 1.0$, $\lambda_2 = 2.0$ and $\eta = 1.0$. Notice the improvement of the Bayesian curvelet-domain separation compared to SRME. Not only are the shallow multiples better removed but we also observe an improved continuity for the late primary arrivals.