Curvelet-domain matched filtering

Felix J. Herrmann, SLIM, EOS-UBC

SUMMARY

Matching seismic wavefields lies at the heart of seismic processing whether one is adaptively subtracting multiples predictions or groundroll. In both cases, the predictions are matched to the actual to-be-separated wavefield components in the observed data. The success of these wavefield matching procedures depends on our ability to (i) control possible overfitting, which may lead to accidental removal of primary energy, (ii) handle data with nonunique dips, and (iii) apply wavefield separation after matching stably. In this paper, we show that the curvelet transform allows us to address these issues by imposing smoothness in phase space, by using their capability to handle conflicting dips, and by leveraging their ability to represent seismic data sparsely.

INTRODUCTION

Matched filtering for the purpose of matching the amplitudes of wavefields prior to subtraction has been an integral part of the seismic data processor's toolbox (see e.g. Verschuur et al., 1992, where matched-filtering is used within Surface-Related Multiple Elimination, SRME). The recent advent of scaling methods for the restoration of migration amplitudes through image-to-remigrated-image matching (see e.g. Guitton, 2004; Herrmann et al., 2008a; Symes, 2008) represents another instance of matched filtering.

Even though matched-filtering procedures have been applied successfully for cases where the data sets differ by only a single short convolutional filter, extensions of this framework to situations where the differences vary non stationarily has been more problematic. In this paper, we propose a method correcting for these nonstationary effects by making the following assumptions: (i) the stationary difference is removed by a global matching procedure, which corresponds to removal of the source/receiver directivity during primary-multiple separation (Herrmann et al., 2008b), and (ii) the remaining nonstationary difference is assumed to vary smoothly in phase space—i.e., the amplitude mismatches are assumed to vary smoothly as a function of position and dip along the coherent wavefronts of the predicted noise component.

THE FORWARD MODEL

With the above assumptions, the nonstationary matching can be represented mathematically by a zero-order pseudo-differential operator (ΨDO), whose action on an arbitrary d-dimensional function is given by

$$\big(Bf\big)(x) = \int_{x\in\mathbb{R}^d} e^{jk\cdot x} b(x,k)\hat{f}(k)\mathrm{d}k \qquad (1)$$

with k the wavenumber vector and $b(x,k)$ a real-valued space- and spatial-frequency dependent filter, known as the symbol. For our application, this operator acts either on shot records or on common-offset panels ($d = 2$) and applies a location, frequency, and dip-dependent scaling. After discretization, this operator models the mismatch by applying a matrix-vector multiplication —i.e.,

$$\mathbf{g} = \boldsymbol{B}\mathbf{f}, \qquad (2)$$

where $\boldsymbol{B}$ is a full positive-definite matrix, implementing the action of the pseudo-differential operator, and $\mathbf{f}$ and $\mathbf{g}$ the two to be matched discretized wavefields. e.g., the total data and predicted (and conventionally matched) multiples, respectively. As shown by Herrmann et al. (2008b), this approach leads to excellent results in primary-multiple on synthethic as well as real data (see also below).

APPROXIMATE FORWARD MODEL

After the appropriate global compensation for the stationary contribution of the matched filter (by e.g. a global matching procedure), the ΨDO in Equation 2 can be considered zero order and hence permits the following diagonal curvelet-domain decomposition (Herrmann et al., 2008a,b),

$$\mathbf{g} \approx \boldsymbol{C}^T \mathrm{diag}\{\mathbf{w}\}\boldsymbol{C}\mathbf{f}, \quad \{w\}_{\mu\in\mathscr{M}} > 0, \qquad (3)$$

with $\boldsymbol{C}$ the 2D discrete curvelet transform (see e.g. Candés et al., 2006) and $\mathbf{w}$ the curvelet-domain scaling vector and $\mathscr{M}$ the index set of curvelet coefficients. Since we are using the curvelet transform based on wrapping, which is a tight frame, we have $\boldsymbol{C}^T\boldsymbol{C} = \boldsymbol{I}\mathbf{d}$ and the transpose, denoted by the symbol ^{T}, equals the pseudo inverse.

In this approximate forward model, for which precise theoretical error estimates exist (Herrmann et al., 2008a), the two wavefields are matched by a simple curvelet-domain scaling. This curvelet-domain scaling applies a location, scale and dip dependent amplitude correction. Since the matrix $\boldsymbol{B}$ is positive-definite, the entries in the scaling vector, $\mathbf{w}$, are positive. This approximate formulation of the forward model is the basis for our curvelet-domain matched filter.

CURVELET-DOMAIN MATCHED FILTERING

Equation 3 lends itself to an inversion for the unknown scaling vector. As both wavefields are known, our formulation minimizes the least-squares mismatch between the two of them. The following issues complicate the estimation of the scaling vector: (i) the undeterminedness of the forward model due to the redundancy of the curvelet transform—i.e., $\boldsymbol{C}\boldsymbol{C}^T$ is rank deficient; (ii) the risk of overfitting the data, which leads to unwanted artifacts such as incorrect amplitude corrections or inadvertent matching of primary energy, and (iii) the positivity requirement for the scaling vector. To address issues (i-ii),

the following augmented system of equations is formed that relates the unknown scaling vector $\mathbf{w}$ to the augmented data vector, $\mathbf{d}$— i.e.,

$$\begin{bmatrix}\mathbf{g}\\ \mathbf{0}\end{bmatrix} = \begin{bmatrix}\boldsymbol{C}^T \operatorname{diag}\{\boldsymbol{C}\check{\mathbf{f}}\}\\ \gamma \boldsymbol{L}\end{bmatrix}\mathbf{w} \tag{4}$$

or $\mathbf{d} = \boldsymbol{F}_\gamma \mathbf{w}$. The scaling vector is found by minimizing the functional

$$J_\gamma(\mathbf{z}) = \frac{1}{2}\|\mathbf{d} - \boldsymbol{F}_\gamma e^{\mathbf{z}}\|_2^2, \tag{5}$$

where the substitution of $\mathbf{w} = e^{\mathbf{z}}$ (with the exponentiation taken elementwise) guarantees positivity (issue (iii)) of the solution (Vogel, 2002). This formulation seeks a solution fitting the vector, $\mathbf{g}$, with a smoothness constraint imposed by the sharpening operator $\boldsymbol{L}$, which for each scale penalize fluctuations amongst neighboring curvelet coefficients in the space and angle directions (see Herrmann et al., 2008a, for a detailed description). The amount of smoothing is controlled by the parameter γ. For increasing γ, there is more smoothness at the expense of overfitting the data (e.g., erroneously fitting the primaries). For a specific γ, the penalty functional in Equation 5 is minimized with respect to the vector $\mathbf{z}$ with the limited-memory BFGS (Nocedal and Wright, 1999) with the gradient

$$\operatorname{grad} J(\mathbf{z}) = \operatorname{diag}\{e^{\mathbf{z}}\}\left[\boldsymbol{F}_\gamma^T\left(\boldsymbol{F}_\gamma e^{\mathbf{z}} - \mathbf{d}\right)\right]. \tag{6}$$

Below, we discuss the application of this curvelet-domain matched filter, each exploiting curvelet domain sparsity.

STABLE APPLICATION OF OUR MATCHED FILTER

Ideally, the scaled multiples yielded by the above nonlinear least-squares problem, $\widetilde{\mathbf{z}} = \arg\min_{\mathbf{z}} J(\mathbf{z})$, could be subtracted from the total data directly. Unfortunately, the presence of noise in seismic imaging (see e.g. Herrmann et al., 2008a) and phase and kinematic errors in primary-multiple separation may interfere, rendering a separation based on the residual alone (as in SRME) ineffective. Following recent work by Saab et al. (2007); Wang et al. (2008), we separate the primaries and multiples by solving the following sparsity-promoting program

$$\begin{aligned}\{\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2\} &= \underset{\mathbf{x}_1,\mathbf{x}_2}{\arg\min}\, \lambda_1\|\mathbf{x}_1\|_{1,\mathbf{w}_1} + \lambda_2\|\mathbf{x}_2\|_{1,\mathbf{w}_2} \\ &+ \|\boldsymbol{A}\mathbf{x}_2 - \mathbf{b}_2\|_2^2 + \eta\|\boldsymbol{A}(\mathbf{x}_1+\mathbf{x}_2) - \mathbf{b}\|_2^2,\end{aligned} \tag{7}$$

where the vectors $\{\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2\}$ represent the estimates for the primaries and multiples, respectively, and where $\boldsymbol{A}$ is the curvelet synthesis matrix, $\{\mathbf{w}_1, \mathbf{w}_2\}$ are positive weights, and $\{\mathbf{b}, \mathbf{b}_2\}$, the total data and the adapted multiple prediction—i.e., $\mathbf{b}_2 = \boldsymbol{A}\operatorname{diag}\{\mathbf{w}\}\boldsymbol{A}^*\check{\mathbf{f}}$ with the symbol * denoting the conjugate transpose. We set the weights in Equation 7 according to $\mathbf{w}_1 = \max\{|\boldsymbol{A}^*\mathbf{b}_2|, \varepsilon\}$ and $\mathbf{w}_2 = \max\{|\boldsymbol{A}^*\mathbf{b}_1|, \varepsilon\}$ with the operations taken elementwise and $\varepsilon > 0$ a small parameter. As detailed in Saab et al. (2007); Wang et al. (2008), the λ's and η are control parameters determining the sparsity of the solution and fits to the total data and the matched multiple prediction.

For appropriately chosen $\lambda_1, \lambda_2, \eta$, and reasonably accurately matched SRME-predicted multiples, the minimization of Equation 7 leads to a separation of the primaries and multiples. To minimize Equation 7, we devise an iterative thresholding algorithm in the spirit of the work by Daubechies et al. (2004). 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

$$\begin{aligned}\mathbf{x}_1^{n+1} &= \boldsymbol{T}_{\frac{\lambda_1\mathbf{w}_1}{2\eta}}\left[\boldsymbol{A}^*\mathbf{b}_2 - \boldsymbol{A}^*\boldsymbol{A}\mathbf{x}_2^n + \boldsymbol{A}^*\mathbf{b}_1 - \boldsymbol{A}^*\boldsymbol{A}\mathbf{x}_1^n + \mathbf{x}_1^n\right] \\ \mathbf{x}_2^{n+1} &= \boldsymbol{T}_{\frac{\lambda_2\mathbf{w}_2}{2(1+\eta)}}\left[\boldsymbol{A}^*\mathbf{b}_2 - \boldsymbol{A}^*\boldsymbol{A}\mathbf{x}_2^n + \mathbf{x}_2^n + \frac{\eta}{\eta+1}\left(\boldsymbol{A}^*\mathbf{b}_1 - \boldsymbol{A}^*\boldsymbol{A}\mathbf{x}_1^n\right)\right],\end{aligned} \tag{8}$$

where $\boldsymbol{T}_{\mathbf{u}} : \mathbb{R}^{|\mathscr{M}|} \mapsto \mathbb{R}^{|\mathscr{M}|}$ is the elementwise soft-thresholding operator—i.e., for each $\mu \in \mathscr{M}$, $T_{u_\mu}(v_\mu) := \operatorname{sgn}(v_\mu)\cdot\max(0, |v_\mu| - |u_\mu|)$. The proposed algorithm provably converges to the minimizer of Equation 7, provided all weights—i.e., all components of the vectors $\mathbf{w}_1$ and $\mathbf{w}_2$, are strictly positive (Daubechies et al., 2004).

Our Bayesian formulation leads to an optimization problem—i.e., the minimization of Equation 7, which involves a combined minimization of the weighted ℓ^1-norms of the coefficient vectors, the ℓ^2-misfit between the matched 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., 2007) since our formulation 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 matched SRME-predicted multiples, the more emphasis we place on the total data misfit. The case $\eta \to \infty$ is analogous to an absolute lack of confidence on the matched SRME-predicted multiples, and thus includes a misfit concerning the total data only. This limiting unrealistic assumption underlaid our earlier formulation (Herrmann et al., 2007). Away from this limit, Equation 7 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. (2007); (iii) the accuracy of the matched 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, we address this situation with the curvelet-domain matched filter described above and in Herrmann et al. (2008b).

SYNTHETIC-DATA EXAMPLE

We consider a shot record from a synthetic line, generated by an acoustic finite-difference code for a velocity model that

consists of a high-velocity layer, which represents salt, surrounded by sedimentary layers and a water bottom that is not completely flat (see Fig. 11 in Herrmann et al., 2007). In Figure 1, the results for optimized single-term SRME are compared to curvelet-domain Bayesian separation with and without our amplitude scaling. Figures 1(a)-1(c) include the total input data with multiples, the SRME-predicted multiples and the "multiple-free" data, respectively. The predicted multiples are the result of conventional matching in a single window. The "multiple-free" data were modeled with an absorbing boundary condition, removing the surface-related multiples. Results for the estimated primaries according to optimized single-term SRME with windowed matching, Bayesian separation and scaled-Bayesian separation are included in Figures 1(d)-1(f). Comparison of these results shows a significant improvement for the primaries computed with the curvelet-domain amplitude scaling, calculated by solving Equation 5 for $\gamma = 0.5$. For this choice of γ, the multiples are not over fitted and the amplitude correction leads to a removal of remnant multiple energy, in particular for the events annotated by the arrows. The value for γ was found experimentally. Finally, notice that the improvement in the estimate for the primaries is due to the combination of curvelet-domain separation and scaling, yielding results that are comparable to the ones expected from multi-term SRME. Even though multi-term SRME, in combination with standard ℓ_2-subtraction, is known to near perfectly remove surface-related multiples for synthetic data, SRME in practice is often only viable for one iteration because field data sets often do not obey assumptions of the model. Therefore, the single-term SRME result in Figure 1(d) can be considered as state of the art.

FIELD DATA EXAMPLE

We test the above-described adaptive separation algorithm by examining real North Sea field dataset. The main purpose of this test is to study the improvement by curvelet-domain matching compared to results obtained with and estimate for the multiples yielded by optimized one-term SRME computed with a windowed-matched filter. This case is relevant for situations where the data quality does not permit iterative SRME or where the cost of multiple iterations of SRME is a concern. In either situation, the predicted multiples will contain amplitude errors, which may give rise to residual multiple energy and dimmed primaries. We show that the proposed scaling by curvelet-domain matched filtering improves the estimation for the primaries as long the curvelet-to-curvelet variations for this scaling are sufficiently controlled by the smoothness constraint. Relaxation of this constraint may leads to overfitting and hence to inadvertent removal of primary energy.

Figure 2(a) contains the common-offset section (at offset 200 m) that we selected from a North Sea field dataset. Estimated primaries according to conventional SRME are plotted in Figure 2(b). Results where ℓ_2-matched filtering in the shot domain (Verschuur and Berkhout, 1997) is replaced by Bayesian thresholding (Saab et al., 2007) in the offset domain, are presented for a single offset in Figure 2(c), without scaling, and in Figure 2(d) with scaling. The scaled result is calculated for $\gamma = 0.3$. Juxtaposing the standard SRME and the curvelet-based results shows a removal of high-frequency clutter, which is in agreement with earlier findings reported in the literature. Moreover, primaries in the deeper part of the section (e.g. near the lower-two arrows in each plot) are much better preserved, compared to the standard-SRME result. Removal of the strong residual multiples in the shallow part, e.g. the first- and second-order water bottom multiples indicated by the arrows around 0.75 and 1.20 s, is particularly exciting. Due to the unbalanced amplitudes of the predicted multiples, both standard SRME and non-adaptive Bayesian thresholding are unable to eliminate these events. Our adaptive method, however, successfully removes these events by virtue of the curvelet-domain scaling. Compared to non-adaptive thresholding, residual multiples are better suppressed, while our adaptive scheme also leads to at least similar, but often even better, overall continuity and amplitude preservation of the estimated primaries. For example, improvements are visible in the lower-left corner of the sections (between offsets $[0-2000]$ m and times $[3.0-3.6]$ s), where low-frequency multiple residuals are better suppressed after curvelet-domain matched filtering (cf. Figure 2(c) and 2(d)), without deterioration of the primary energy. Finally, observe the improved recovery of primary energy at the lower arrow in Figure2(d), compared to the primary in Figure 2(c).

RELATION TO RECENT WORK AND EXTENSIONS

In a recent conference proceeding, Neelamani et al. (2008) propose a elementwise curvelet-domain adaptive subtraction method that is based on the complex-valued curvelet transform. Their method involves the solution of the following constrained optimization problem for each individual curvelet coefficient—i.e,

$$\begin{aligned} \min_{a_\mu,\,\phi_\mu} \; & |\{\boldsymbol{C}_c\mathbf{g}\}_\mu - a_\mu \exp(j\phi_\mu)\{\boldsymbol{C}_c\check{\mathbf{f}}\}_\mu| \\ \text{subject to} \quad & a_\mu^{min} \le a_\mu \le a_\mu^{max}, \\ & \phi_\mu^{min} \le \phi_\mu \le \phi_\mu^{max}, \\ & \text{for} \quad \mu \in \mathscr{M}. \end{aligned} \tag{9}$$

In this expression, $\boldsymbol{C}_c$ is the complex curvelet transform, and a_μ, ϕ_μ are the elementwise amplitude and phase factors that minimize the difference between the complex curvelet coefficients of the total data and of the predicted multiples. To prevent overfitting, these factors are bound between elementwise minimal and maximal values—i.e., the a_μ^{min}, a_μ^{max} and ϕ_μ^{min}, ϕ_μ^{max}.

Even though this is an elegant and computationally efficient formalism that yields good results, the method only matches *locally*—i.e., for each element of the curvelet vector separately without curvelet-to-curvelet control—while our method matches the complete curvelet coefficient vector by solving a *global* nonlinear optimization problem where curvelet-to-curvelet variations in the corrections are controlled by a smoothness constraint (cf. Equation 5). This smoothness constraint is controlled by a single Lagrange multiplier γ (cf. Equation 5) that is related to our insistence on modeling the mismatch between

the actual and predicted multiples by a ΨDO with a smooth and positive symbol.

However, in practice the above assumption may be violated since multiple predictions and actual multiples generally differ by kinematic shifts. Allowing for shifts as part of a matching procedure would solve this problem but extreme care should be taken because SRME-based multiple predications are solidly grounded by the wave equation, and applying significant shifts entails a departure that may not be justifiable. However, as rightfully pointed out by Neelamani et al. (2008), complex curvelet transforms (see also Herrmann, 2005, for a discussion on phase rotated wavelets) allow for small phase rotations remedying the lack of shift invariance of the real-valued curvelet transform. So far, we used this property after amplitude matching during our Bayesian separation (Saab et al., 2007; Wang et al., 2008). This Bayes separation is also based on the complex curvelet transform and involves another *global* optimization problem now constrained by the matched multiple predictions and the total data mismatch (cf. Equation 7).

So far, the adaptive part of our formulation was aimed at adapting the amplitudes, leaving the kinematic errors to the second Bayes stage of the separation. During the workshop we plan to report whether incorporation of an extension towards a complex-valued matching procedure will further improve our results.

DISCUSSION AND CONCLUSIONS

In this paper, we presented a comprehensive method to compute the coefficients of a curvelet-domain matched filter and to apply the estimated filter through a Bayes separation method where sparsity on the to-be-separated signal components is promoted. The stylized example with the conflicting dips underlined the importance of imposing phase-space smoothness while the field data example showed the merit of our method in the context of adaptive primary-multiple separation.

ACKNOWLEDGMENTS

The author would like to thank Eric Verschuur for providing us with the SRME-primaries. I would also like to thank D. Wang, R. Saab and O. Yilmaz for their contributions to this work and the authors of CurveLab for making their codes available. Norsk Hydro is thanked for making the field dataset available. This work was in part financially supported by the NSERC Discovery Grant (22R81254) of F.J.H. and CRD Grant DNOISE (334810-05).

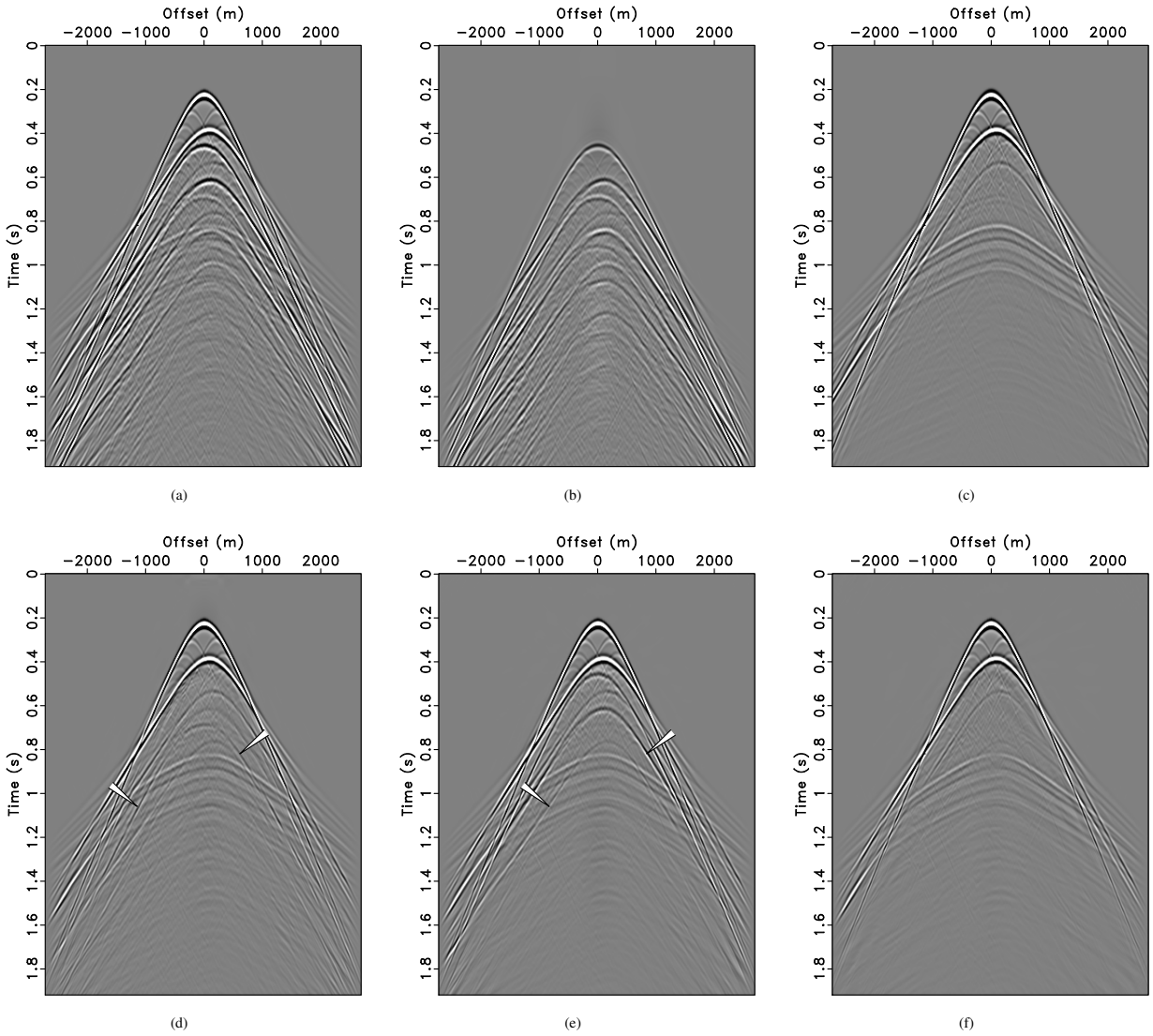

Figure 1: Primary-multiple separation on a synthetic shot record. **(a)** The total data, $\mathbf{p}$, including primaries and multiples. **(b)** Single-term SRME-predicted multiples wavelet-matched within a global window ($\check{\mathbf{s}}_2$). **(c)** Reference surface-related multiple-free data modeled with an absorbing boundary condition. **(d)** Estimate for the primaries, yielded by optimized one-term SRME computed with a windowed-matched filter. **(e)** Estimate for the primaries, computed by Bayesian iterative thresholding with a threshold defined by $\mathbf{t} = |\boldsymbol{C}\check{\mathbf{s}}_2|$. **(f)** The same as **(e)** but now for the scaled threshold, i.e., $\mathbf{t} = |\mathrm{diag}\{\widetilde{\mathbf{w}}\}\boldsymbol{C}\check{\mathbf{s}}_2|$ (with $\gamma = 0.5$). Notice the improvement for the scaled estimate for the primaries, compared to the primaries yielded by SRME in **(d)** and by the Bayesian separation without scaling in **(e)**.

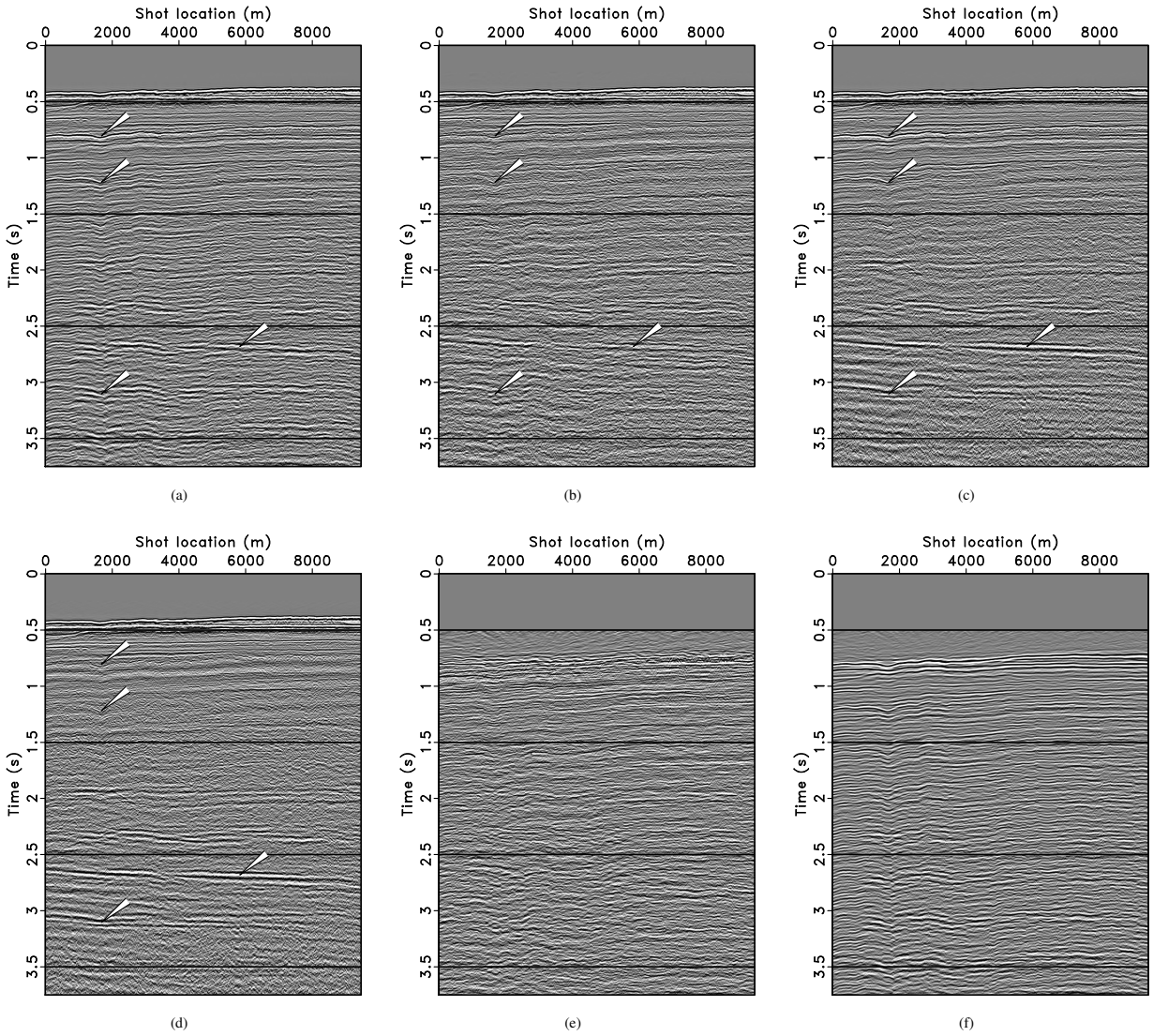

Figure 2: Adaptive curvelet-domain primary-multiple separation on real data. **(a)** Near-offset (200 m) section for the total data plotted with automatic-gain control. **(b)** Estimate for the primaries, yielded by optimized one-term SRME computed with a windowed-matched filter. **(c)** Estimate for the primaries, computed by Bayesian iterative thresholding with a threshold defined by $\mathbf{t} = |\boldsymbol{C}\check{\mathbf{b}}_2|$ with $\mathbf{b}_2$ the predicted multiples. **(d)** The same as **(c)** but now for the scaled (for $\gamma = 0.3$) threshold, i.e., $\mathbf{t} = |\mathrm{diag}\{\widetilde{\mathbf{w}}\}\boldsymbol{C}\mathbf{b}_2|$ with $\widetilde{\mathbf{w}} = \exp(\widetilde{\mathbf{z}})$. **(e)** The difference between SRME and matched filter. **(f)** Difference between the total data and the matched predicted primaries. Notice the improvement for the scaled estimate for the primaries, compared to the primaries yielded by SRME in **(b)** and by the Bayesian separation without scaling in **(c)**.

REFERENCES

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

Daubechies, I., M. Defrise, and C. De Mol, 2004, An iterative thresholding algorithm for linear inverse problems with a sparsity constraints: Comm. Pure Appl. Math., **57**, 1413–1457.

Guitton, A., 2004, Amplitude and kinematic corrections of migrated images for nonunitary imaging operators: Geophysics, **69**, 1017–1024.

Herrmann, F. J., 2005, Seismic deconvolution by atomic decomposition: a parametric approach with sparseness constraints: Integrated Computer-Aided Engineering, **12**, no. 1, 69–90.

Herrmann, F. J., U. Boeniger, and D. J. Verschuur, 2007, Nonlinear primary-multiple separation with directional curvelet frames: Geophysical Journal International, **170**, no. 2, 781–799.

Herrmann, F. J., P. P. Moghaddam, and C. C. Stolk, 2008a, Sparsity- and continuity-promoting seismic imaging with curvelet frames: Journal of Applied and Computational Harmonic Analysis, **24**, 150–173. (doi:10.1016/j.acha.2007.06.007).

Herrmann, F. J., D. Wang, and D. J. Verschuur, 2008b, Adaptive curvelet-domain primary-multiple separation: Geophysics, **73**, no. 3, A17–A21.

Neelamani, R., A. Baumstein, and W. S. Ross, 2008, Adaptive subtraction using complex curvelet transforms: Presented at the EAGE 70th Conference & Exhibition, Rome, G048.

Nocedal, J. and S. J. Wright, 1999, Numerical optimization: Springer.

Saab, R., D. Wang, O. Yilmaz, and F. Herrmann, 2007, Curvelet-based primary-multiple separation from a bayesian perspective: Presented at the SEG International Exposition and 77th Annual Meeting.

Symes, W. W., 2008, Approximate linearized inversion by optimal scaling of prestack depth migration: Geophysics, **73**, R23–R35. (10.1190/1.2836323).

Verschuur, D. J. and A. J. Berkhout, 1997, Estimation of multiple scattering by iterative inversion, part II: practical aspects and examples: Geophysics, **62**, 1596–1611.

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

Vogel, C., 2002, Computational Methods for Inverse Problems: SIAM.

Wang, D., R. Saab, O. Yilmaz, and F. J. Herrmann, 2008, Bayesian-signal separation by sparsity promotion: application to primary-multiple separation: Technical report, UBC Earth and Ocean Sciences Department. (TR-2008-1. accepted for publication in GEOPHYSICS).