

STANFORD EXPLORATION PROJECT

*Patricia Berge, James Berryman, Biondo Biondi, Brian Bonner, Morgan Brown,
Henri Calandra, Jon Claerbout, Robert Clapp, Sean Crawley, Sergey Fomel,
Douglas Gratwick, Antoine Guitton, Carmen Mora, Marie Prucha, James Rickett,
Daniel Rosales, Paul Sava, and Louis Vaillant*

Report Number 103, April 2000

`../../paul1/fig/salt-comp.pdf`

Copyright © 2001

by the Board of Trustees of the Leland Stanford Junior University

Stanford, California 94305

*Copying permitted for all internal purposes of the Sponsors of
Stanford Exploration Project*

Preface

The electronic version of this report¹ makes the included programs and applications available to the reader. The markings [ER], [CR], and [NR], are promises by the author about the reproducibility of each figure result.

Reproducibility is a way of organizing computational research, which allows both the author and the reader of a publication to verify the reported results at a later time. Reproducibility facilitates the transfer of knowledge within SEP and between SEP and its sponsors.

ER denotes Easily Reproducible and are the results of a processing described in the paper. The author claims that you can reproduce such a figure from the programs, parameters, and makefiles included in the electronic document. The data must either be included in the electronic distribution, or be easily available (e.g. SEG-EAGE data sets) to SEP and non-SEP researchers. The data may also be available in the SEP data library, which can be viewed at http://sepwww.stanford.edu/public/docs/sepdata/lib/toc_html/. We assume you have a UNIX workstation with Fortran, Fortran90, C, X-Windows system and the software downloadable from our website (SEP makerules, SEPlib, and, to properly reproduce the documents, the SEP latex package). Before the publication of the electronic document, someone other than the author tests the author's claim by destroying and rebuilding all ER figures. Some ER figures may not be reproducible by outsiders because they depend on data sets that are too large to distribute, or data that we do not have permission to redistribute but are in the SEP data library.

CR denotes Conditional Reproducibility. The author certifies that the commands are in place to reproduce the figure if certain resources are available. SEP staff have only attempted to make sure that the makefile rules exist and the source codes referenced are provided. The primary reasons for the CR designation is that the processing requires 20 minutes or more, or commercial packages such as Matlab or Mathematica.

M denotes a figure that may be viewed as a movie in the web version of the report. A movie may be either ER or CR.

NR denotes Non-Reproducible. This class of figure is considered non-reproducible. SEP discourages authors from flagging their figures as NR except for artist drawings, scanings, etc.

Our testing is currently limited to IRIX 6.5 and LINUX 2.1 (using the Portland Group Fortran90 compiler), but the code should be portable to other architectures. Reader's suggestions are welcome. For more information on reproducing SEP's electronic documents, please visit <http://sepwww.stanford.edu/redoc/>.

Jon Claerbout, Biondo Biondi, Robert Clapp, Sergey Fomel, and Marie Prucha

¹<http://sepwww.stanford.edu/private/docs/sep103>

SEP-103 — TABLE OF CONTENTS

Velocity

<i>Clapp R. G. and Biondi B. L.</i> , Tau tomography with steering filters: 2-D field data example	1
<i>Sava P. and Biondi B.</i> , Wave-equation migration velocity analysis: Episode II	19
<i>Clapp R. G.</i> , Regularizing tomography with non-stationary filters	47
<i>Rickett J.</i> , Traveltime sensitivity kernels: Banana-doughnuts or just plain bananas?	61
<i>Rosales D.</i> , Research interest: Improving the velocity model	69

Imaging

<i>Biondi B.</i> , Stable wide-angle Fourier-finite difference downward extrapolation of 3-D wavefields	79
<i>Rickett J.</i> , Efficient 3-D wavefield extrapolation with Fourier finite-differences and helical boundary conditions	103
<i>Prucha M. L., Clapp R. G., and Biondi B.</i> , Seismic image regularization in the reflection angle domain	109
<i>Sava P. and Fomel S.</i> , Angle-gathers by Fourier Transform	119
<i>Claerbout J.</i> , Spectral factorization of 2-D reflection seismic data	131
<i>Vaillant L. and Calandra H.</i> , Common-azimuth migration and Kirchhoff migration for 3-D prestack imaging: A comparison on North Sea data	139
<i>Sava P.</i> , Variable-velocity prestack Stolt residual migration with application to a North Sea dataset	147
<i>Vaillant L. and Biondi B.</i> , Accuracy of common-azimuth migration approximations	157
<i>Gratwick D.</i> , Sub-salt imaging of a 2-D elastic synthetic model, using prestack, split-step, wave equation migration	169

Coherent noise

<i>Guitten A.</i> , Prestack multiple attenuation using the hyperbolic Radon transform	181
<i>Clapp R. G. and Brown M.</i> , $(t - x)$ domain, pattern-based multiple separation	201
<i>Claerbout J. and Fomel S.</i> , Spitz makes a better assumption for the signal PEF	211

<i>Brown M. and Claerbout J.</i> , Ground roll and the Radial Trace Transform – revisited	219
---	-----

Optimization

<i>Fomel S.</i> , Seismic data interpolation with the offset continuation equation	237
<i>Guitton A.</i> , Huber solver versus IRLS algorithm for quasi L1 inversion	255
<i>Claerbout J. and Fomel S.</i> , Speculations on contouring sparse data: Gaussian curvature . . .	271
<i>Guitton A.</i> , Implementation of a nonlinear solver for minimizing the Huber norm	281

Helix filtering

<i>Fomel S.</i> , Helical preconditioning and splines in tension	289
<i>Crawley S.</i> , Patching and micropatching in seismic data interpolation	301
<i>Claerbout J.</i> , The Burg Method on a Helix?	311
<i>Rickett J.</i> , Data alignment with non-stationary shaping filters	315
<i>Claerbout J.</i> , Random lines in a plane	325
<i>Claerbout J.</i> , Test case for PEF estimation with sparse data	329

Reservoir characterization

<i>Berryman J. G., Berge P. A., and Bonner B. P.</i> , Transformation of seismic velocity data to extract porosity and saturation values for rocks	331
<i>Mora C. and Biondi B.</i> , Estimation of AVO attributes sensitivity to velocity uncertainty using forward modeling: a progress report	349

SEP article published or in press, 2000-01	367
SEP phone directory	369
Research personnel	371
SEP sponsors for 2000-01	377

Tau tomography with steering filters: 2-D field data example

Robert G. Clapp and Biondo L. Biondi¹

ABSTRACT

Common reflection point (CRP) gathers are usually parameterized in terms of offset. For tomography, a more convenient parameterization is in terms of reflection angle. Reflection angle CRP gathers can be constructed using wave equation migration. By performing tomography in vertical travel-time (τ), focusing velocity, rather than some combination of focusing and mapping velocity, can be estimated. By using anisotropic preconditioning oriented along bedding planes, the solution can be guided towards a geologically reasonable model. The benefits of the tomography method are shown on a 2-D line from a 3-D marine dataset.

INTRODUCTION

Tomography is inherently non-linear, therefore a standard technique is to linearize the problem by assuming a stationary ray field (Stork and Clayton, 1991). Unfortunately, we must still deal with the coupled relationship between reflector position and velocity (Al-Chalabi, 1997; Tieman, 1995). As a result, the back projection operator must attempt to handle both repositioning of the reflector *and* updating the velocity model (van Trier, 1990). The resulting back projection operator is sensitive to our current guess at velocity and reflector position. In vertical travelttime space, reflector movement is significantly less. Biondi et al. (1998b) showed that by reformulating the problem in this space, complex velocity functions could be obtained more quickly and are better resolved.

In ray-based migration velocity analysis, Kirchhoff migration is normally used to construct CRP gathers and residual moveout. Unfortunately, conventional implementation of Kirchhoff methods have trouble handling complex wave behavior. Wave equation methods are an attractive alternative to Kirchhoff. Clayton and Stolt (1981) and later Prucha et al. (1999) showed that wave equation methods can form CRP gathers in terms of reflection angle.

Tomography problems are also often under-determined. By adding an additional model regularization term to our objective function (Toldi, 1985) we can stabilize the inversion. In theory, this regularization term should be the inverse model covariance matrix (Tarantola, 1987). Clapp et al. (1998a) constructed a series of small plane wave annihilators, called steer-

¹**email:** bob@sep.stanford.edu,biondo@sep.stanford.edu

ing filters, using *a priori* information to produce changes to our velocity model that were more geologically reasonable.

In this paper we show how to apply steering filters to smooth the slowness, rather than the change in slowness. We use wave equation CRP gathers as the basis for ray based tomography. We apply the technique on a 2-D line taken from a 3-D marine dataset from the North Sea. We show that we can obtain a velocity model that improves the focusing of the data while being geologically reasonable.

REVIEW

Following the methodology of Clapp and Biondi (1999) we will begin by considering a regularized tomography problem. We will linearize around an initial slowness estimate and find a linear operator in the vertical traveltimes domain \mathbf{T} between our change in slowness $\Delta\mathbf{s}$ and our change in traveltimes $\Delta\mathbf{t}$. We will write a set of fitting goals,

$$\begin{aligned}\Delta\mathbf{t} &\approx \mathbf{T}\Delta\mathbf{s} \\ \mathbf{0} &\approx \epsilon\mathbf{A}\Delta\mathbf{s},\end{aligned}\tag{1}$$

where \mathbf{A} is our steering filter operator and ϵ is a Lagrange multiplier. However, these fitting goals don't accurately describe what we really want. Our steering filters are based on our desired slowness rather than change of slowness. With this fact in mind, we can rewrite our second fitting goal as:

$$\begin{aligned}\mathbf{0} &\approx \mathbf{A}(\mathbf{s}_0 + \Delta\mathbf{s}) \\ -\epsilon\mathbf{A}\mathbf{s}_0 &\approx \epsilon\mathbf{A}\Delta\mathbf{s}.\end{aligned}\tag{2}$$

$$\tag{3}$$

Our second fitting goal can not be strictly defined as regularization but we can do a preconditioning substitution (Fomel et al., 1997):

$$\begin{aligned}\Delta\mathbf{t} &\approx \mathbf{T}\mathbf{A}^{-1}\mathbf{p} \\ -\epsilon\mathbf{A}\mathbf{s}_0 &\approx \epsilon\mathbf{I}\mathbf{p}.\end{aligned}\tag{4}$$

Wave equation angle gathers

When doing velocity analysis, general practice is to measure moveout from a relatively sparse set of CRP gathers. Kirchhoff depth migration is the preferred construction method because it can produce the sparse set of CRP gathers without needing to image the entire volume. In addition, if our Green's function table is constructed correctly, Kirchhoff methods do not suffer from the velocity approximations needed by frequency domain methods. Kirchhoff methods also have some deficiencies. The most glaring weakness of Kirchhoff methods is the difficulty in constructing the Green's function table. To construct an accurate Green's

function table we must account for, and weight correctly, the multiple arrivals that occur in complex geology. Calculating and accounting for multiple arrivals adds significantly to both coding complexity and memory requirements. As a result, a single arrival is often all that is used. Eikonal methods (Vidale, 1990; van Trier and Symes, 1991; Podvin and Lecomte, 1991; Fomel, 1997) can efficiently produce first arrivals, but in areas of complex geology the first arrival is not always the most important arrival (Audebert et al., 1997). Nichols (1994) proposed a band-limited method that gave the maximum amplitude arrival, but the method is computationally impractical in 3-D. As a result, people usually go to expensive ray based methods but still face the difficult tasks of choosing the most important arrival and correctly and efficiently interpolate the traveltimes field (Sava and Biondi, 1997).

The most computationally attractive alternative to Kirchhoff methods is frequency domain downward continuation. Downward continuation has its own weaknesses. Its primary weakness is speed. Downward continuation can not be target oriented, so full volume imaging is required. In addition, frequency domains methods in their purest form can not handle lateral variations in velocity. By migrating with multiple velocities and applying a space domain correction to the wavefield, they can do a fairly good job handling lateral variations (this migration is normally referred to as PSPI, Phase-shift plus interpolation)(Ristow and Ruhl, 1993). Finally, downward continuation focuses the wavefield towards zero offset, making conventional moveout analysis impossible.

We can create CRP gathers where moveout analysis is possible by changing our imaging condition (Clayton and Stolt, 1981; Prucha et al., 1999). Given a wave-field we follow the normal procedure of downward continuing the data and extracting the image at the surface $z = 0$ and zero time. Instead of extracting the image at zero offset, we note that reflection angle θ can be evaluated by the differential equation:

$$\tan \theta = -\frac{\partial z}{\partial x_h} \quad (5)$$

where z is the depth, x_h is half-offset.

The topic of this paper is not migration, but tomography. The tomography method could be applied with either Kirchhoff or PSPI. For us, PSPI proved to be a more attractive choice. A 2-D and 3-D PSPI algorithm was already available, where a Kirchhoff approach would have required the coding of the migration algorithm along with a suitable traveltimes computation method.

Characterizing moveout errors

Tomography requires us to provide moveout errors. It is unreasonable to hand pick every reflector at every CRP gather in 2-D and inconceivable in 3-D. As a result, people have tried to find alternate methods to pick moveout errors. Clark et al. (1996) used a neural network approach to pick CRP gathers and many people have suggested seeding-based approaches to pick the gathers. Both approaches describe complicated moveouts, but they suffer from cycle skipping and have problems in areas where the S/N ratio is not very high. An alternative

approach is to characterize the moveout in CRP gathers by a single parameter (Etgen, 1990; Biondi, 1990). A single parameter is a much more robust estimator. It requires less human involvement (less picking and/or QA is necessary) and is less sensitive to signal to noise problems.

At early iterations a single parameter is especially valuable. All that can be resolved at early iterations are gross features. A single parameter can capture these where picking the entire CRP gather is likely to cause the inversion to be overwhelmed small features that are not resolvable at early iterations. When we were close to the correct velocity allowing freedom in moveout behavior is desirable and beneficial.

For the tomography problem we will begin with a migrated image d at a depth z , angle θ , at CRP location x . For estimating the residual moveout in the CRP gathers by calculating semblance s in terms of some curvature parameter γ ,

$$s(z, c, x) = \frac{[\sum_{\theta} d(z + \theta\gamma^2, \theta, x)]^2}{n(z, x) \sum_{j=0} d(z + \theta\gamma^2, x)^2}, \quad (6)$$

where $n(z, x)$ is the number non-zero samples summed over for each semblance calculation.

For this dataset hand picking the semblance along each reflector would not be too tedious, but in 3-D it would quickly become so. As a result, we wanted to come up with a simple way for the computer to do most of the work. One option would be to just pick the maximum semblance at each location, but we can get an unrealistic, high spatial wavenumber behavior for $\gamma(x)$. When doing convention semblance analysis we are confronted with a similar problem, that picking the maximum semblance at each time could result in an unreasonable velocity function. Clapp et al. (1998b) proposed a method to avoid hand picking that still led to a reasonable velocity model. We can adapt that work by starting with the maximum curvature value at each CRP γ_{\max} , the semblance at the maximum curvature value \mathbf{W} , and a derivative operator \mathbf{D} . We can find a smooth curvature function $\boldsymbol{\gamma}$ by setting up a simple set of equations

$$\begin{aligned} \mathbf{0} &\approx \mathbf{W}(\boldsymbol{\gamma}_{\max} - \boldsymbol{\gamma}) \\ \mathbf{0} &\approx \epsilon \mathbf{D}\boldsymbol{\gamma}. \end{aligned} \quad (7)$$

By increasing ϵ we get smoother $\boldsymbol{\gamma}$ values while small ϵ values honor more our maximum semblance picks.

Endpoints, edge effects, and errors

To set up our tomography problem we need to cover some final details. We can convert our semblance picks back into a Δz shift by applying

$$\Delta z = \gamma\theta^2. \quad (8)$$

We then note that our tomography problem is set up for time rather than depth errors, To convert our depth error to a time error we multiplying by RMS slowness s_{rms} of. If we use

a straight ray geometric assumption, we can approximate the time error at a given offset by multiplying by $\cos \theta$ and $\cos \phi$ (where ϕ is the geologic dip) obtaining as our final relationship

$$\Delta t = \frac{2\gamma\theta^2 \cos \theta \cos \phi}{v(z, \mathbf{x})}. \quad (9)$$

In constructing our raypaths we benefit from having our CRP gathers in terms of angle. If errors were in terms of offset we would have to either

- shoot rays from every source and receiver location, and find ray pairs that obey Snell's law at the position on the reflector imaged at the offset dictated by the ray pair
- or shoot rays from our CRP locations along where our reflector is imaged at every offset and then interpolate the ray field to our source and receiver locations.

Both options require significant additional ray-tracing in 2-D, and even more in 3-D. In addition, we are always faced with the tradeoff of how much should we interpolate our rays versus how many additional rays should we shoot.

With our moveout errors in terms of angle we only need to shoot a single ray-pair up from our imaging point at the angle α and β ,

$$\begin{aligned} \alpha &= \phi + \theta \\ \alpha &= \phi - \theta \end{aligned} \quad (10)$$

where θ is one-half the aperture angle, ϕ is the geologic dip (Figure 1). If the rays emerge at surface locations corresponding to an offset and CMP location inside our acquisition geometry we have a valid ray pair.

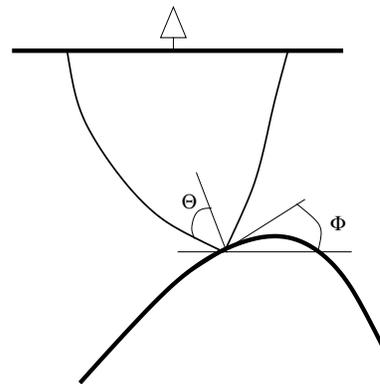


Figure 1: How the takeoff angle for a ray-pair are defined. bob1-sketch
[NR]

DATA

The tomography method described in the preceding section is suited for a particular class of problems. Using rays and picking reflectors requires that the dataset it is applied to be

relatively clean with strong, fairly continuous reflectors. Formulating the tomography in tau is universally advantageous, but is most advantageous in environments where we have a velocity inversion. Finally, steering filters are most effective when velocity follows structural dip, ruling out most Gulf of Mexico environments.

The North Sea dataset provided to SEP by Elf Aquitaine (Vaillant and Sava, 1999; Sava and Fomel, 2000; Fomel, 2000) meets all of these criteria. The data is very clean (Figure 2) with strong reflectors that are generally continuous. The data contains a chalk layer which causes a velocity inversion above the salt dome. The initial velocity model (Figure 3) was created by Elf using the S.M.A.R.T² method (Jacobs et al., 1992; Ehinger and Lailly, 1995). The velocity structure shows typical North Sea behavior with velocity following structural layers.

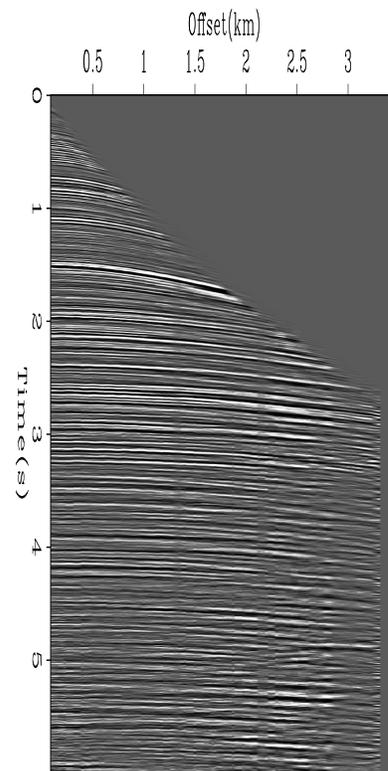
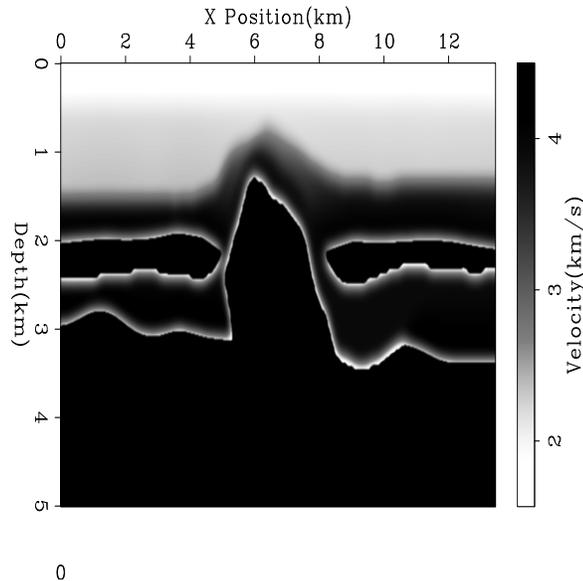


Figure 2: A shot gather from the Elf Aquitaine dataset. `bob1-elf-shot` [ER]

The dataset is a 3-D marine acquired using three cables with 100 meter spacing and geophones every 25 meters. In this paper we will be dealing with a 2-D subset of the 3-D dataset. The line was chosen coincide with a 2-D synthetic dataset (Prucha et al., 1998; Malcotti and Biondi, 1998). The subset was created by forming partial stacking and then applying Azimuth Moveout (AMO) (Biondi et al., 1998a) to partial stacked CMP gathers.

²Sequential Migration-Aided Reflection Tomography - KIM (Kinematic Inversion Methods), IFP consortium

Figure 3: The initial velocity model for the 2-D line. Note how velocity follows structural dip and the velocity reversal in the shale layer above the salt dome. `bob1-elf.vel0` [ER]



INITIAL ERRORS

We began by migrating the data with the velocity in Figure 3. As Figure 4 shows we have some artifacts in the right portion of the migrated image due to the sharp boundary in the initial velocity model. The salt is poorly defined. We don't have a coherent reflection for the top of the salt, we have very little of the salt edges, and have a non-continuous salt bottom. Reflections to the left of the salt body do not continue all the way to the salt edge. In addition, we have little resolution below the fault at 10 km. If we look at the angle gathers for the initial migration, Figure 5, we can see significant moveout errors. Using the initial migrated image we chose 11 reflectors to perform tomography with (Figure 6). To constrain the upper portion of the model we chose the water bottom reflection and two reflectors above the salt. We picked the salt top and salt bottom and three reflectors on both sides of the salt body. We then performed moveout analysis using equation (6). We selected the semblance at each reflector, Figure 7, and found a smooth curve using fitting goals (9). Overlaid on Figure 7 are the smooth picks used in our back projection operator. The top two reflectors have almost no moveout errors. The remaining reflectors all have some residual moveout errors that our tomography can attempt to resolve.

Building the steering filters

To construct the steering filters we used the nine non-salt reflectors shown in Figure 6. We calculated the dip along each reflector and then smoothly interpolated between the reflectors (Figure 8). With the picked reflectors we have very few rays passing through the salt. As a result the smoothing fitting goal (the second goal in (4)) dominates. Instead of constant or little variation in the salt we see dramatic changes. To avoid this geologically unreasonable behavior we did not allow the velocity to change in the region delineated by our two salt reflectors.

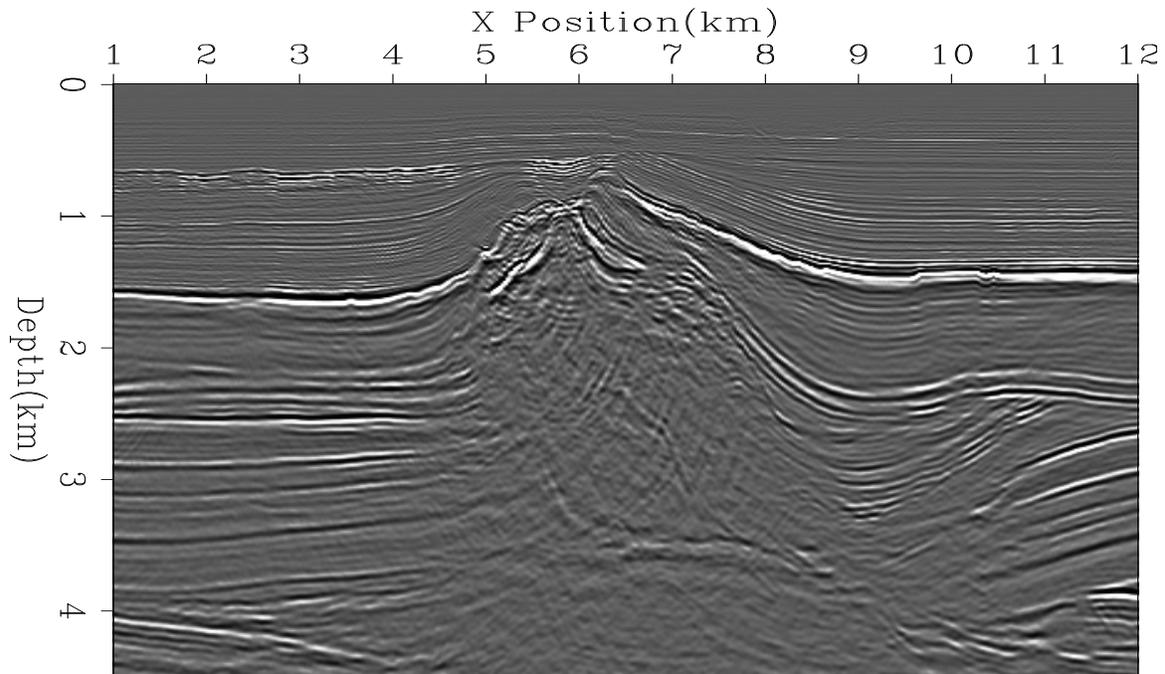


Figure 4: Migration result using the velocity from Figure 3 `bob1-elf.mig.vel0` [CR]

RESULTS

We performed two non-linear iteration of tomography. Figure 9 shows our velocity model. Note that our velocity still follows structural dip. This is especially obvious around the fault at 10 km and above the top of the salt. If we examine the same CRP gathers, Figure 11, we see that we have done a better job flattening the reflectors and have higher frequency events. If we look at the semblance along the reflectors, Figure 10, we see that have significantly less moveout. The true test is the final migrated image. If we look at the final migrated image we can see better continuity of the reflectors and generally higher frequency image. The final migration image (Figure 12) is quite similar to our initial image (Figure 4). If we take a closer look at the top of the salt body (Figure 13) we can begin to see some differences. At location ‘A’ and ‘C’ we have done a better job defining the salt boundary. At ‘B’ the reflector is sharper and has a more realistic shape. If we look at the side of the salt dome, Figure 14, we can see more improvements. The reflectors are more continuous ‘A’ and extend closer to the salt boundary ‘B’. We also have done a better job recovering the bottom salt reflector, ‘C’.

CONCLUSIONS

In this paper we applied our tomography methodology to a 2-D dataset. We showed how to construct a complex steering filter operator based on reflector geometry and *a priori* knowledge of the acoustic properties of the layers. The final migration result was significantly improved over the initial migration showing flatter angle gathers and overall crisper image.

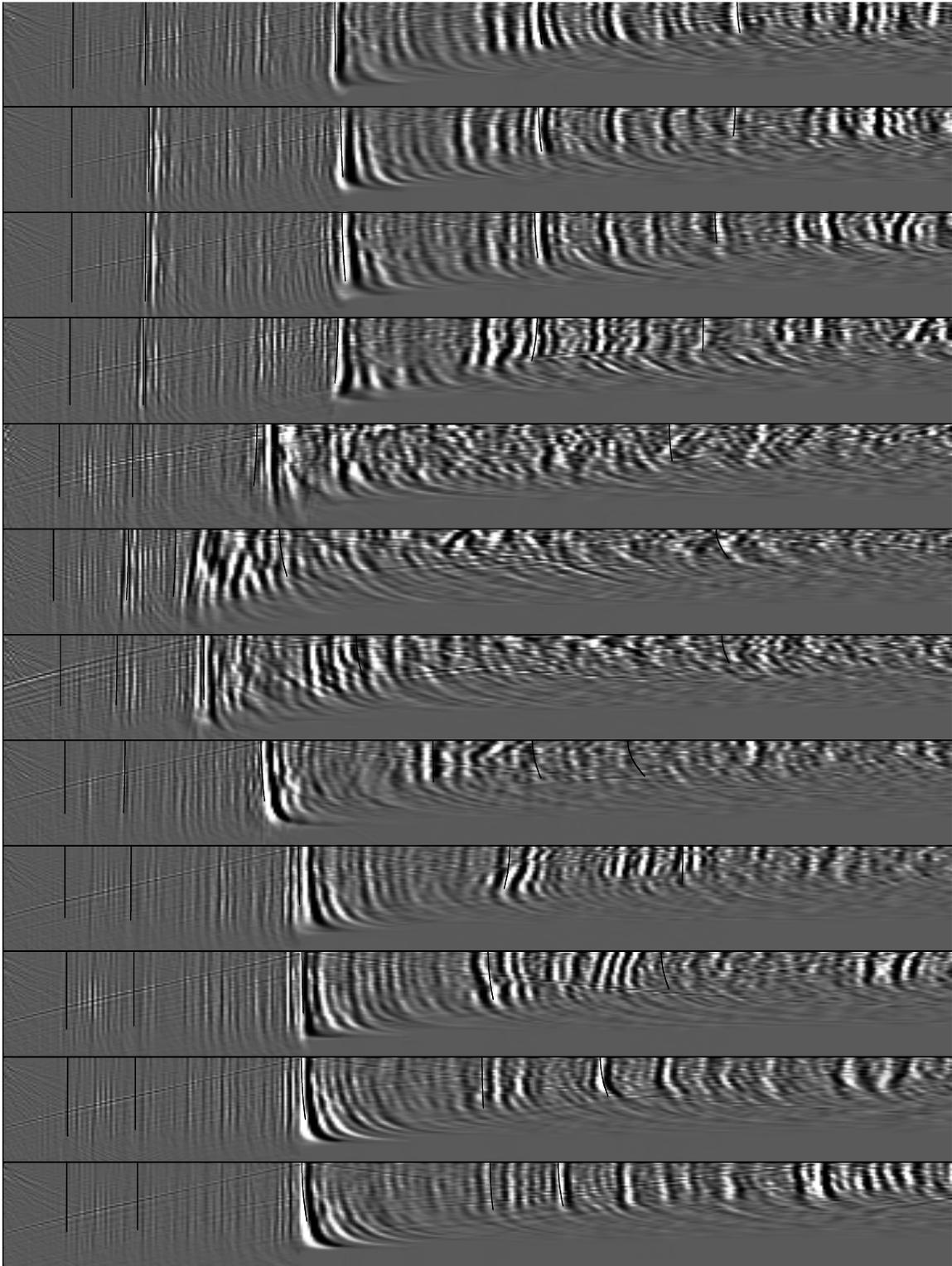


Figure 5: CRP gathers from the initial velocity overlaid by the moveout picks for the reflectors used in the first iteration of tomography. `bob1-overlays.vel0` [CR]

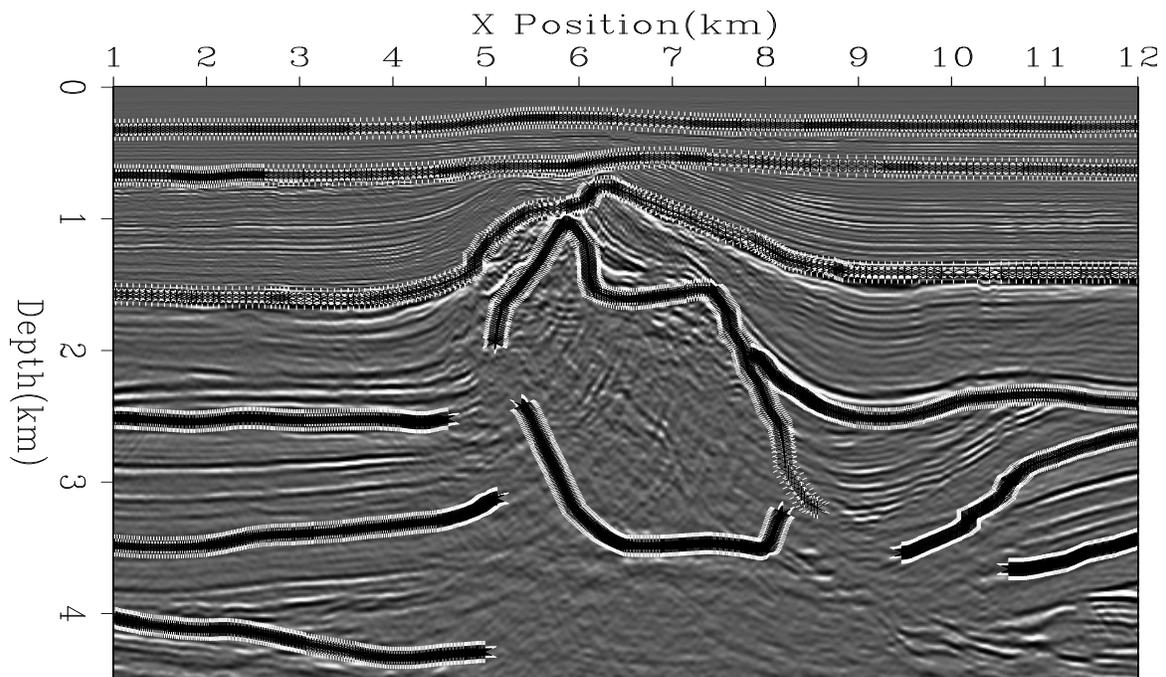


Figure 6: Initial migration with picked reflectors overlaid. `bob1-overlays` [CR]

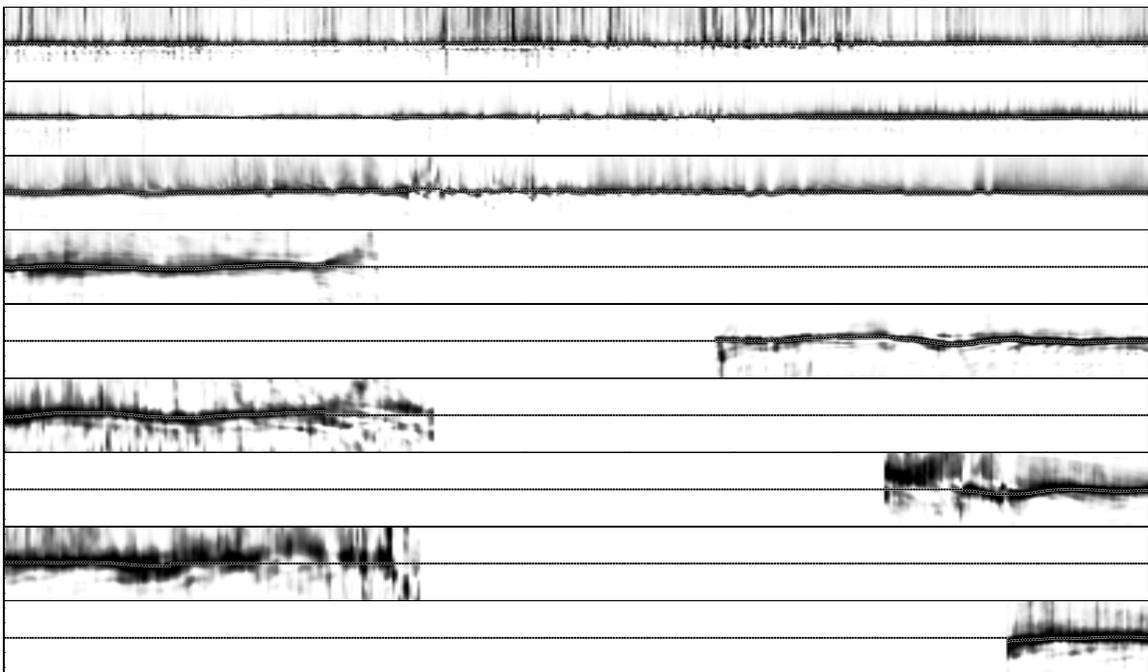


Figure 7: Semblance panels from nine of the reflectors used in tomography. Note that all but the top two still have significant residual moveout. `bob1-elf-sem-b-mig0-ref` [CR]

Figure 8: The dip field used for the first iteration of tomography.
bob1-amp-vel0 [ER]

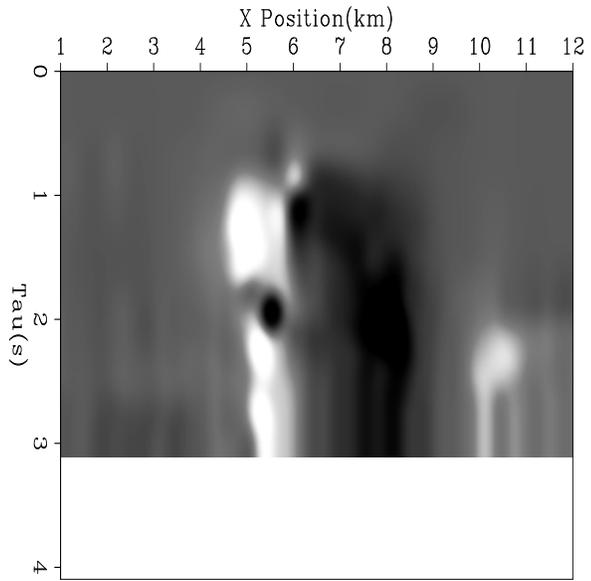
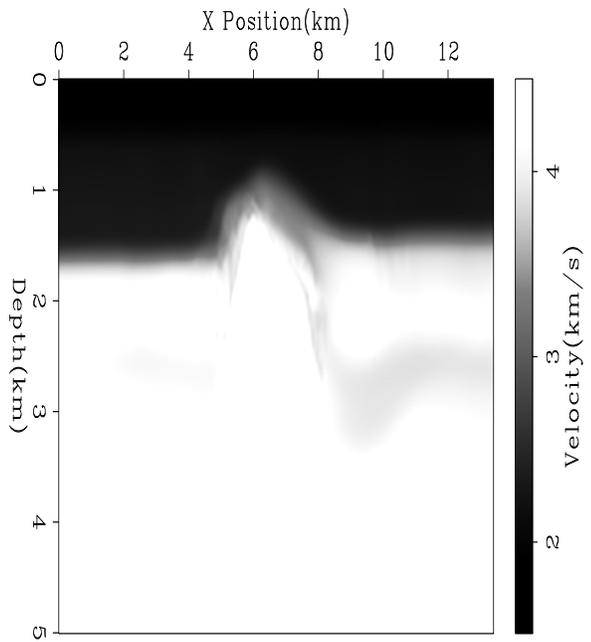


Figure 9: Final velocity
bob1-vel-final [CR]



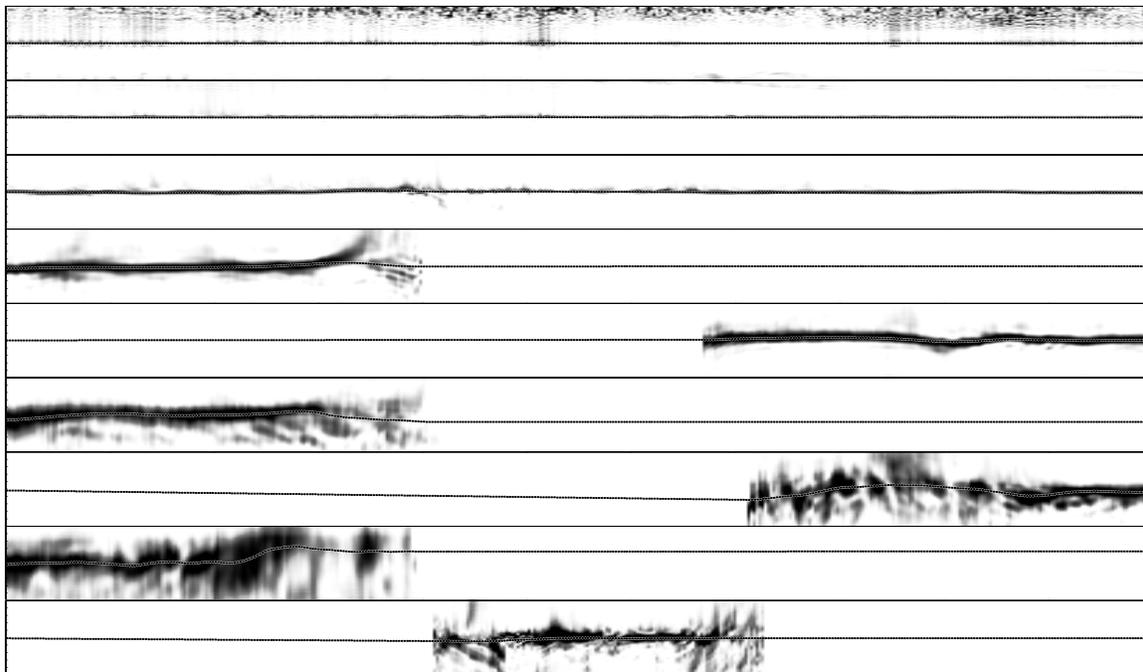


Figure 10: Semblance along the reflectors using final velocity model. `bob1-elf-semb-mig-final-ref` [CR]

We hypothesize that much of the remaining moveout is due to 3-D effects, not resolvable by a 2-D algorithm.

ACKNOWLEDGEMENTS

We would like to thank Elf for providing the data.

REFERENCES

- Al-Chalabi, M., 1997, Parameter nonuniqueness in velocity versus depth functions: *Geophysics*, **62**, no. 03, 970–979.
- Audebert, F., Nichols, D., Rekdal, T., Biondi, B., Lumley, D. E., and Urdaneta, H., 1997, Imaging complex geologic structure with single-arrival kirchhoff prestack depth migration: *Geophysics*, **62**, no. 05, 1533–1543.
- Biondi, B., Fomel, S., and Chemingui, N., 1998a, Azimuth moveout for 3-d prestack imaging: *Geophysics*, **63**, no. 02, 574–588.
- Biondi, B. L., Clapp, R. G., Fomel, S. B., and Alkhalifah, T. A., 1998b, Robust reflection tomography in the time domain: 68th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1847–1850.

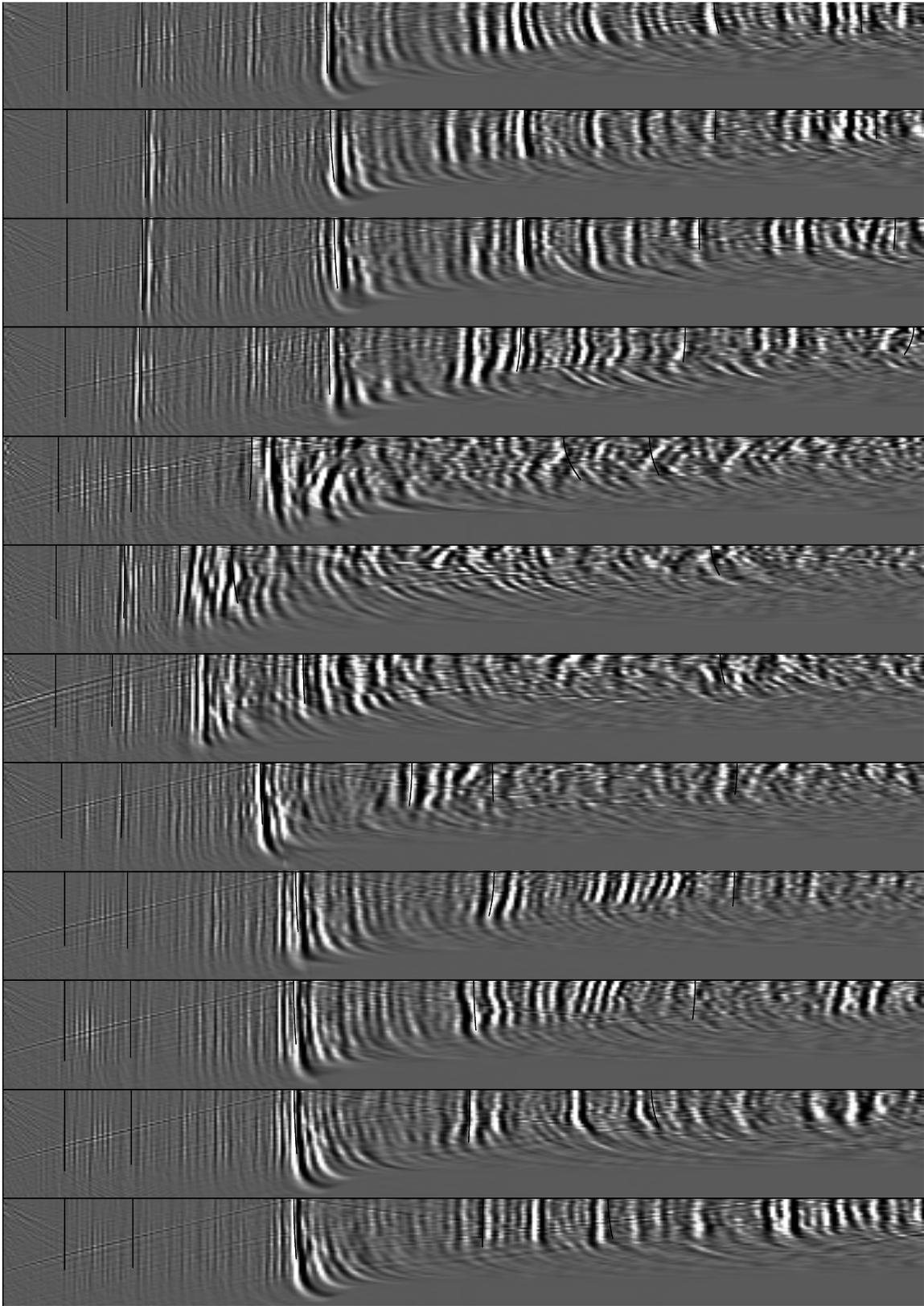


Figure 11: CRP gathers after final migration. `bob1-overlays.vel-final` [CR]

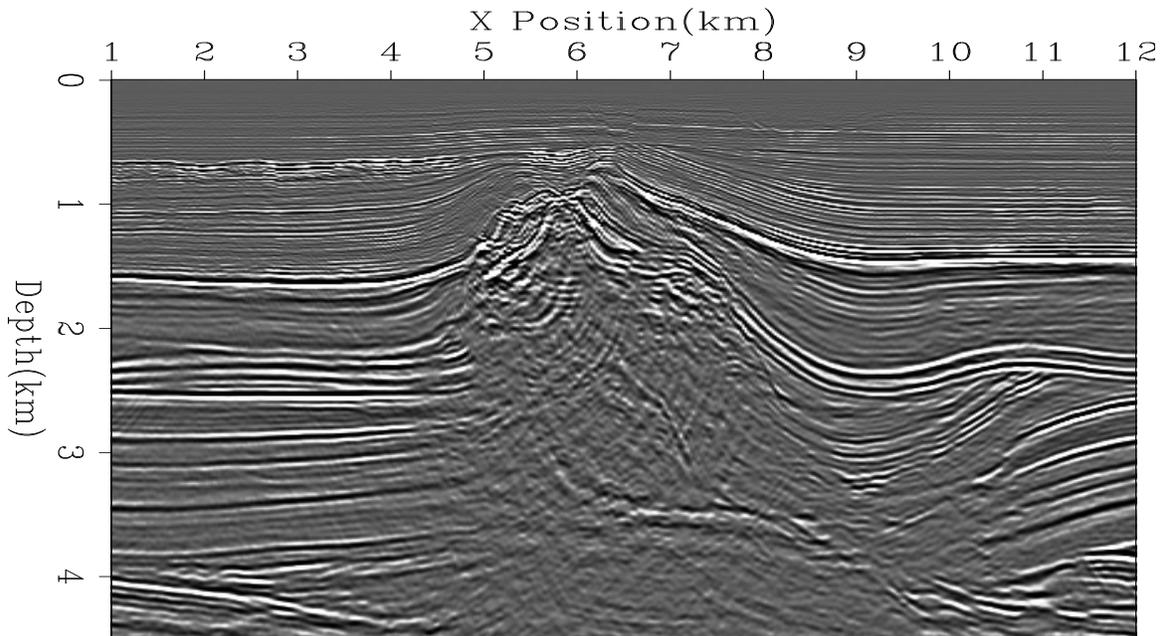


Figure 12: Final migrated image. `bob1-mig-final` [CR]

Biondi, B., 1990, Velocity analysis using beam stacks: Ph.D. thesis, Stanford University.

Clapp, R. G., and Biondi, B., 1999, Preconditioning tau tomography with geologic constraints: SEP-100, 35–50.

Clapp, R. G., Biondi, B. L., Fomel, S. B., and Claerbout, J. F., 1998a, Regularizing velocity estimation using geologic dip information: 68th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1851–1854.

Clapp, R. G., Sava, P., and Claerbout, J. F., 1998b, Interval velocity estimation with a null-space: SEP-97, 147–156.

Clark, G. A., Glinsky, M. E., Devi, K. R. S., Robinson, J. H., Cheng, P. K. Z., and Ford, G. E., 1996, Automatic event picking in pre-stack migrated gathers using a probabilistic neural network: 66th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 735–738.

Clayton, R. W., and Stolt, R. H., 1981, A born-wkby inversion method for acoustic reflection data: Geophysics, **46**, no. 11, 1559–1567.

Ehinger, A., and Lailly, P., 1995, Velocity model determination by the SMART method, part 1: Theory: 65th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 739–742.

Etgen, J., 1990, Residual prestack migration and interval velocity estimation: Ph.D. thesis, Stanford University.

Fomel, S., Clapp, R., and Claerbout, J., 1997, Missing data interpolation by recursive filter preconditioning: SEP-95, 15–25.

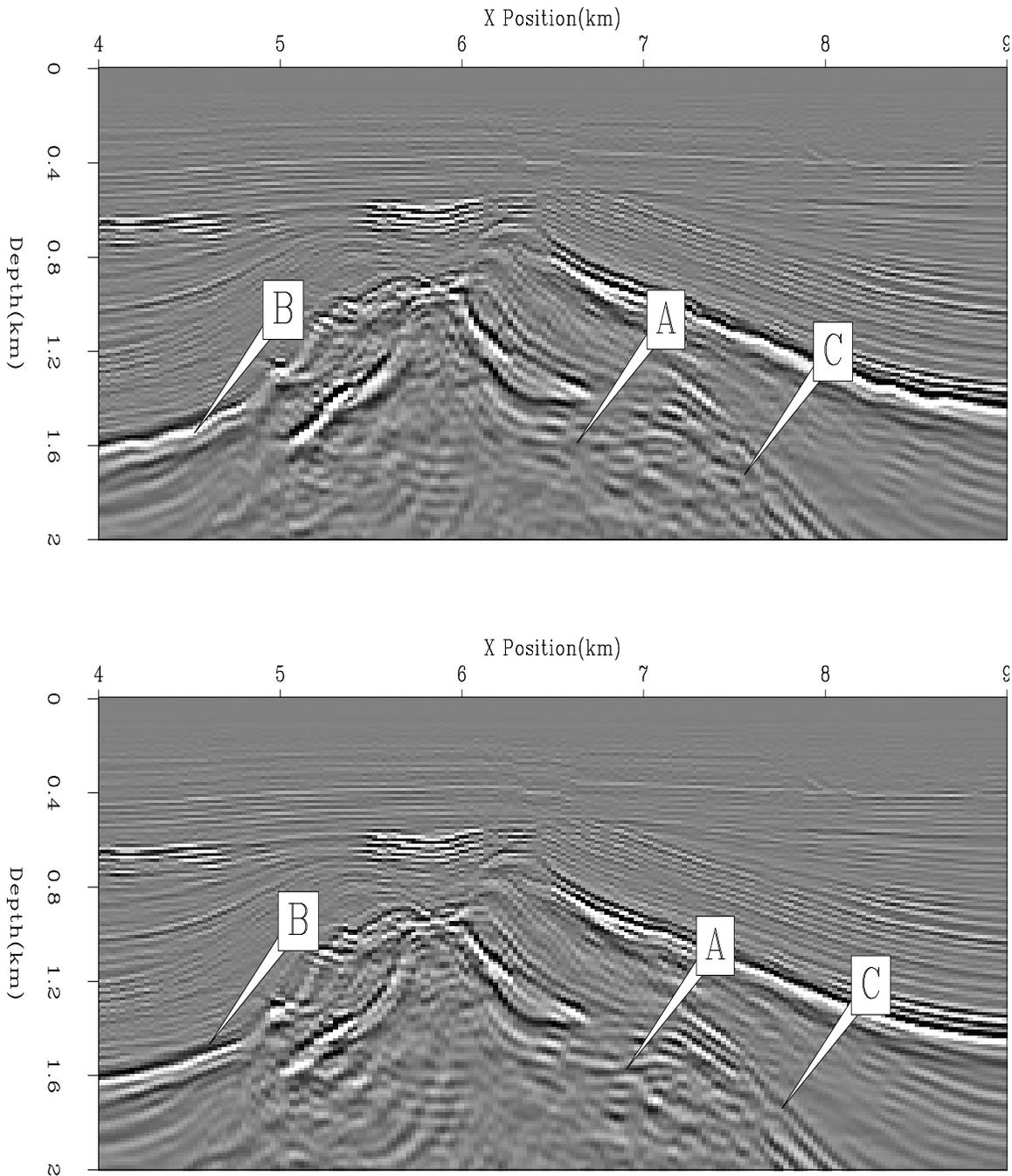


Figure 13: Comparison at the top of the salt dome. At location 'A' and 'C' we have done a better job defining the salt boundary. At 'B' we can see that the reflector is sharper and has a more realistic shape. `bob1-mig-top` [CR]

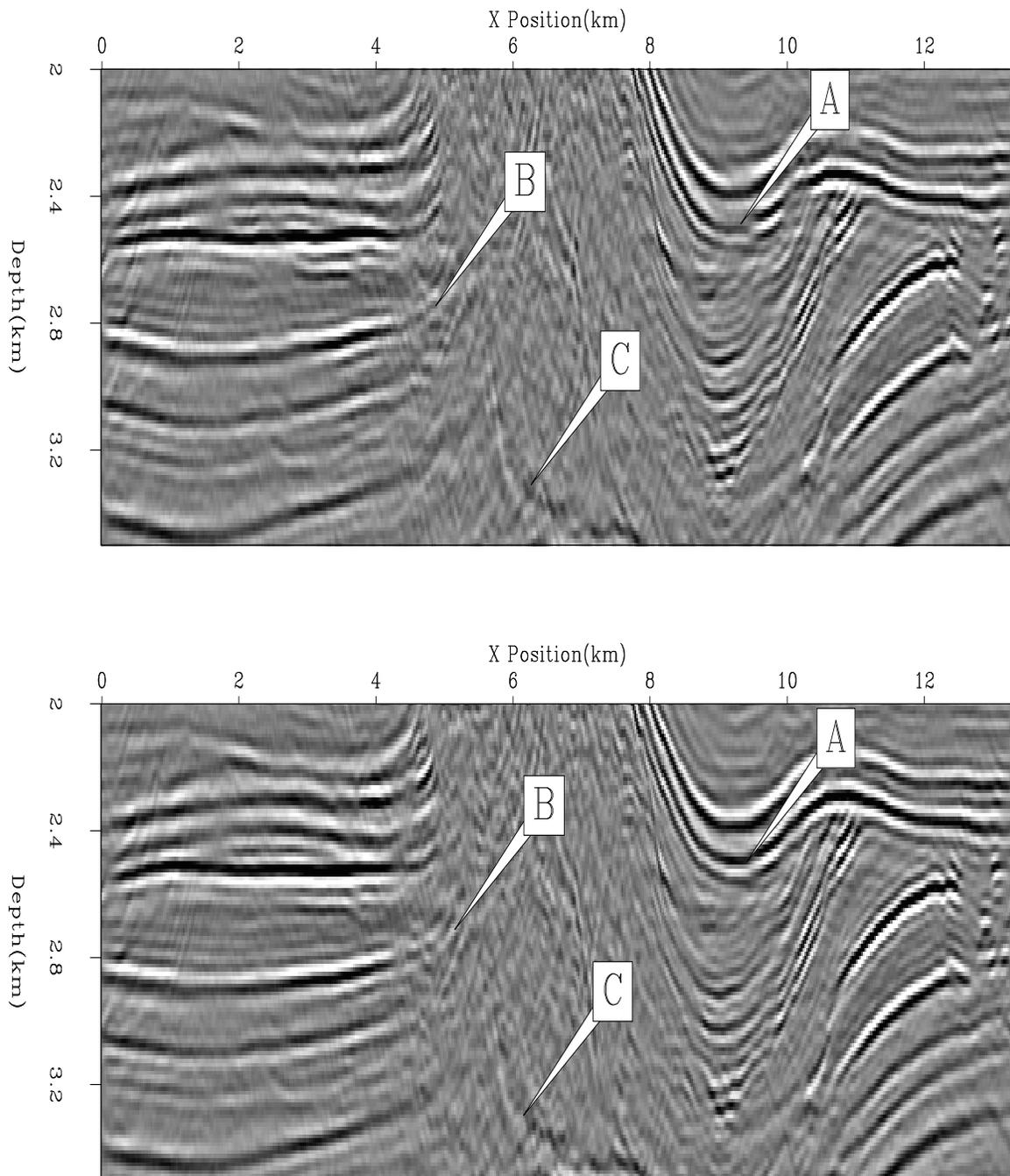


Figure 14: Comparison of the top of the salt dome. The reflectors are more continuous 'A' and extend closer to the salt boundary 'B'. We also have done a better job recovering the bottom salt reflector, 'C'. `bob1-mig-side` [CR]

- Fomel, S., 1997, A variational formulation of the fast marching eikonal solver: SEP-95, 127–147.
- Fomel, S., 2000, Helical preconditioning and splines in tension: SEP-103, 289–301.
- Jacobs, J. A. C., Delprat-Jannaud, F., Ehinger, A., and Lailly, P., 1992, Sequential migration-aided reflection tomography: A tool for imaging complex structures: 62nd Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1054–1057.
- Malcotti, H., and Biondi, B., 1998, Accurate linear interpolation in the extended split-step migration: SEP-97, 61–72.
- Nichols, D. E., 1994, Imaging complex structures using band-limited Green's functions: SEP-81.
- Podvin, P., and Lecomte, I., 1991, Finite difference computation of traveltimes in very contrasted velocity models: A massively parallel approach and its associated tools: *Geophysical Journal International*, **105**, 271–284.
- Prucha, M. L., Clapp, R. G., and Biondi, B. L., 1998, Imaging under the edges of salt bodies: Analysis of an Elf North Sea dataset: SEP-97, 35–44.
- Prucha, M. L., Biondi, B. L., and Symes, W. W., 1999, Angle-domain common image gathers by wave-equation migration: 69th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 824–827.
- Ristow, D., and Ruhl, T., 1993, Extended split-step migration by cascading phase shift and finite difference operators: 63rd Annual Internat. Mtg., Soc. Expl. Geophys., Expanding Abstracts, 986–989.
- Sava, P., and Biondi, B., 1997, Multivalued traveltimes interpolation: SEP-95, 115–126.
- Sava, P., and Fomel, S., 2000, Angle-gathers by Fourier Transform: SEP-103, 119–130.
- Stork, C., and Clayton, R. W., 1991, An implementation of tomographic velocity analysis: *Geophysics*, **56**, no. 04, 483–495.
- Tarantola, A., 1987, *Inverse problem theory*: Elsevier.
- Tieman, H. J., 1995, Migration velocity analysis: Accounting for the effects of lateral velocity variations: *Geophysics*, **60**, no. 01, 164–175.
- Toldi, J., 1985, Velocity analysis without picking: SEP-43.
- Vaillant, L., and Sava, P., 1999, Common-azimuth migration of a North Sea dataset: SEP-102, 1–14.
- van Trier, J., and Symes, W. W., 1991, Upwind finite-difference calculation of traveltimes: *Geophysics*, **56**, no. 6, 812–821.

van Trier, J., 1990, Tomographic determination of structural velocities from depth migrated seismic data: Ph.D. thesis, Stanford University.

Vidale, J. E., 1990, Finite-difference calculation of traveltimes in three dimensions: *Geophysics*, **55**, no. 5, 521–526.

Wave-equation migration velocity analysis: Episode II

*Paul Sava and Biondo Biondi*¹

ABSTRACT

We elaborate the main points of the wave-equation migration velocity analysis method introduced in a previous report. We analyze its strengths and limitations, and illustrate them using a synthetic example. The inversion results confirm our original expectations, especially with regard to stability and robustness. The main difficulty in recovering a complete velocity perturbation is related to the Born approximation, which limits the amount of residual migration improvement on the seismic images at any one given step.

INTRODUCTION

Migration and velocity analysis are intertwined components of seismic imaging. Depth migration cannot be done without knowledge about the velocity model, nor can velocity analysis be done without reasonably accurate seismic images. Furthermore, because migration and migration velocity analysis are complementary objects, it makes intuitive sense to require that both are based on the same physics of wave propagation, namely either on the band-limited wave-equation or its high-frequency asymptotic.

Our previous publication (Biondi and Sava, 1999) introduced the concept of wave-equation migration velocity analysis (WEMVA) as the natural counterpart of wave-equation migration. The objective of our method is to exploit in the velocity analysis context the main strengths of processing based on the wave-equation: accuracy, multipathing, and stability.

In addition to its wave-equation nature, another important concept exploited by our method is that of velocity analysis by image enhancement, in contrast to other methods that also use wave-equation techniques but aim at fitting the recorded data. Our method operates by recursively and simultaneously improving both the migrated image and the velocity model. The advantage of doing so is that the users have an opportunity to quality control the results at every step.

In this paper, we briefly review the theory of WEMVA and show a complete synthetic example in which we start with simulated data, create an image with an approximate velocity model, and then improve both the velocity and the image using WEMVA.

¹email: paul@sep.stanford.edu, biondo@sep.stanford.edu

The WEMVA puzzle relies on several key ingredients, among which some of the most important are image enhancement by residual migration, image quality control in the angle-domain and inversion that uses model constraints. All these are briefly discussed in this paper and are demonstrated with our example.

THEORY

In wave-equation migration, data recorded at the surface are recursively downward continued in depth to generate the wavefield at every depth level (Figure 1). Once the complete wavefield is generated, we can apply an imaging condition to obtain the image at every depth level. In practice, however, downward continuation and imaging are merged into one step, so that we do not need to store the complete downward continued wavefield for all frequencies.

The imaging process described in the preceding paragraph requires that we make an assumption about the slowness model. We label this slowness the background model; therefore, the wavefield and image generated using this model are also naturally labeled as the background wavefield and image.

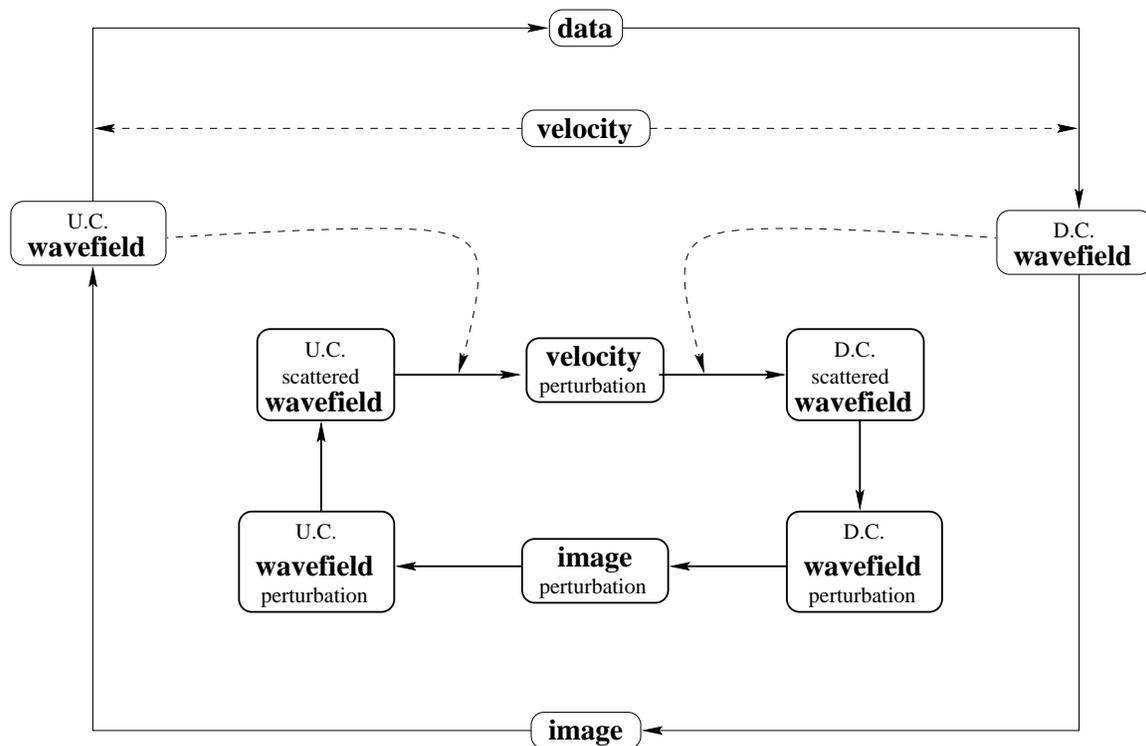


Figure 1: The chart of wave-equation migration velocity analysis technique. D.C. is a label for objects involved in the downward continuation branch, while U.C. is a label for the similar objects of the upward-continuation branch. `paul3-core` [NR]

Wave-equation modeling, the adjoint of migration, proceeds by upward-continuing the data from the highest depth and by adding contributions at every depth level. Eventually, we

reach the surface and reconstruct the wavefield at the surface, which is equivalent, but not identical, to the recorded data.

A perturbation in the slowness model, excited by the background wavefield, generates a scattered wavefield. Like the background wavefield, this is downward continued, creating a wavefield perturbation, and is imaged, creating an image perturbation. In practice, we can either compute the background wavefield at the same time we compute the scattered wavefield, or we can compute and store it before, one frequency at a time.

As for the background, the adjoint process takes us from the image perturbation to an upward-continued wavefield from which we can derive at every depth level a scattered wavefield and, furthermore, a slowness perturbation.

This is the core theoretical result of the wave-equation migration velocity analysis technique: we can relate a perturbation in the slowness model to a perturbation in the migrated image. This enables us to improve the slowness model at the same time we improve the migrated image. This observation formally materializes a well-known concept: migration and velocity analysis are fundamentally related and inseparable from each other (Gray, 1998). This concept is not particular to our method, although the power of WEMVA lies in that it formally eliminates the distinction between velocity analysis and migration and aims at one single goal of image improvement.

Our previous publication (Biondi and Sava, 1999) established the theoretical framework of the method. We have presented the mathematical foundations of the method and a few simple experiments. Here we go one step further and present a complete example starting with the data at the surface and ending with an improved slowness model. Figure 2 shows a flowchart of our method, that consists of the following major loop:

1. Wave-equation migration,
2. image enhancement by residual migration in the angle domain,
3. velocity inversion,
4. slowness update and re-migration.

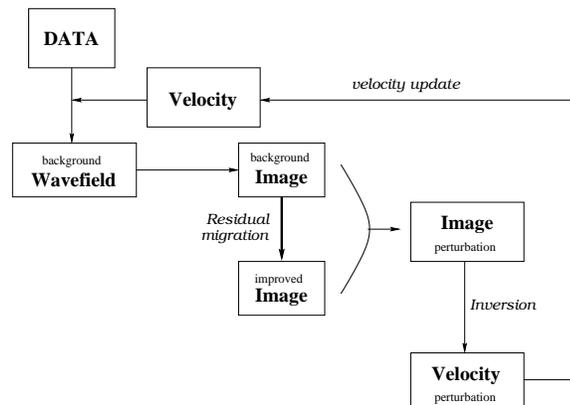


Figure 2: A flowchart of wave-equation migration and WEMVA.

paul3-flow [NR]

Residual migration

Image enhancement is one of the most critical operations of the entire technique. Several methods can in theory be used to enhance the migrated image, for example, residual move-out and residual migration. The major difference between the two is that residual moveout does not allow energy to move among neighboring horizontal locations, while residual migration does. Sava (1999) shows how prestack Stolt residual migration can be employed to enhance migrated images described as angle-domain common-image gathers. In this paper, we use residual migration for image enhancement, but concentrate more on the results related to inversion and slowness update.

Slowness inversion

Once we have created an image perturbation , we can invert for the corresponding perturbation in slowness. Mathematically, this amounts to solving an optimization problem like (Claerbout, 1999)

$$\begin{aligned} \mathcal{L}\Delta\mathbf{s} &\approx \Delta\mathbf{R} \\ \epsilon\mathcal{A}\Delta\mathbf{s} &\approx 0, \end{aligned} \quad (1)$$

where

- \mathcal{L} is a data-fitting operator, mainly composed of a scattering and a downward continuation operator, but which also incorporates the background wavefield (Biondi and Sava, 1999),
- \mathcal{A} is a model-styling operator, either an isotropic Laplacian or an anisotropic steering-filter (Clapp and Biondi, 1998), which imposes smoothing on the model,
- $\Delta\mathbf{s}$ and $\Delta\mathbf{R}$ are respectively the slowness perturbation (the model) and the image perturbation (the data),
- ϵ is a scalar parameter controlling the weight of each of the individual goals.

To speed-up the inversion procedure, we can precondition the model in Equation 1 and solve the system

$$\begin{aligned} \mathcal{L}\mathcal{A}^{-1}\Delta\mathbf{p} &\approx \Delta\mathbf{R} \\ \epsilon\Delta\mathbf{p} &\approx 0, \end{aligned} \quad (2)$$

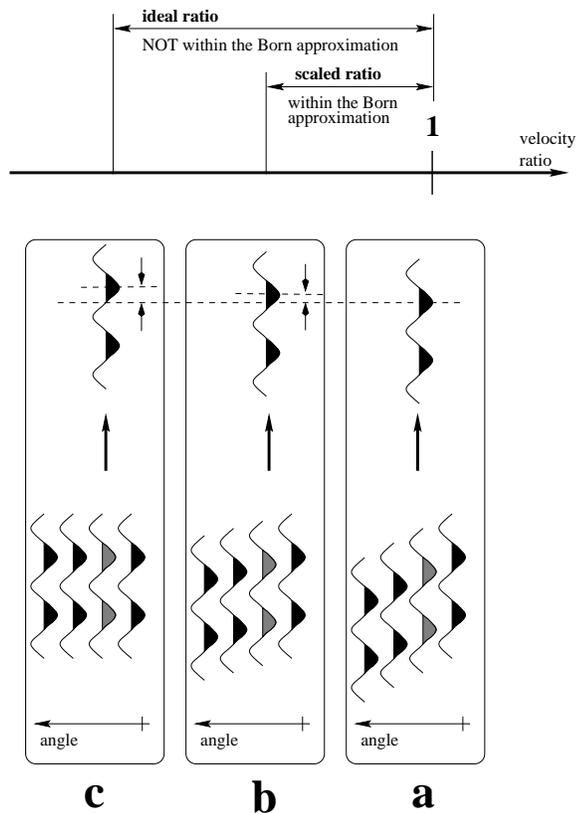
where $\Delta\mathbf{p} = \mathcal{A}\Delta\mathbf{s}$ is the preconditioned model variable.

Slowness update and the Born approximation

We generate the image perturbation by taking the difference between a reference image and an improved version of it. An underlying assumption of the inversion process is that we are within the Born approximation, which requires that the wavelets with which we operate are not more than $\pi/4$ out of phase. What happens if this condition is not fulfilled? In other words, what can we do if the enhanced image is so different from the original image that the wavelets we subtract are not in-phase anymore?

A possible answer to this question, although not necessarily the only one, is that we need to be conservative at the time we generate the improved image. For this, we can scale the velocity-ratio parameter surface (Sava, 1999), which controls the amount of enhancement in the image closer to unity, that is, closer to the original image (Figure 3).

Figure 3: An illustration of velocity ratio scaling which makes the Born approximation valid. (a) is a gather extracted from the original image ($\gamma = 1$). (c) is the same gather after residual migration with the correct ratio. (b) is the same gather after residual migration with a scaled ratio. Although the ideal image is represented by (c), we cannot use this image because the wavelets are not within the Born approximation. Instead, we use (b), which is within the Born approximation and indicates the same direction of image improvement as the optimal ratio. paul3-ratscale [NR]



The shortcoming of this procedure is that we reduce the magnitude of image perturbation, although we preserve a more important parameter – its direction. The scaled-down restored image falls within the limits of the Born approximation with respect to the original; therefore, we can safely invert for the slowness perturbation. However, the slowness perturbation we obtain depends on the scaling we have done on the images, although it has the correct direction.

Next we need to scale the slowness perturbation back up to the value corresponding to the correct velocity ratio measured from the residual migrated images. So, how do we do this? A possible solution is to run a line search using the slowness perturbation we have inverted, with

the goal of maximizing the energy of the migrated image. Mathematically, this goal can be expressed as

$$\max_{\alpha} \|\mathcal{L}(\mathbf{s} + \alpha \Delta \mathbf{s})\|, \quad (3)$$

where α is the scaling factor for the model, and \mathbf{s} is the background slowness model.

EXAMPLE

We exemplify our velocity analysis technique using a synthetic model. The reflectivity, depicted in Figure 4, shows a structure with steeply dipping reflectors and an reverse fault. To complicate matters further, we set the slowness so that the fractured layer has higher velocity than the surrounding area (Figure 5). Although it doesn't seem difficult, this model is complicated by the severe distortion of the wavefield that propagates under the faulted region. Such a geological setting is common in overthrust regions and also around salt bodies.

Figure 4: The reflectivity model.

`paul3-mod1.refl` [CR]

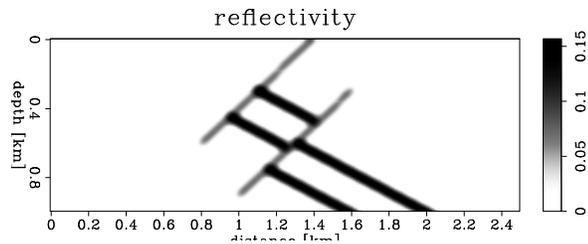
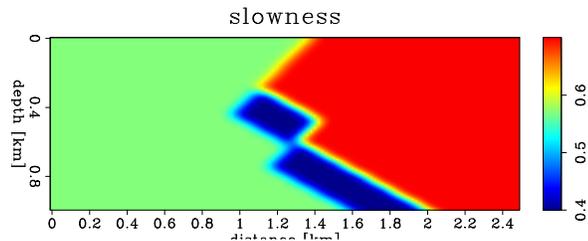


Figure 5: The true slowness model.

`paul3-mod1.Slow.0` [CR]



We begin by generating a synthetic dataset. Based on our simulated acquisition, defined by a half-offset of 2.0 km, we expect to achieve a reasonably good angular coverage of the region of interest, which is not deep.

The goal of this test is to start the imaging process by assuming that we don't know much about the slowness of the fast layer and to use our technique to fill the missing slowness. We run two separate tests: a simpler test when we assume that we know the geometry and slowness of the upper part of the model, and a more complex test when we do not assume anything about the slowness in the layer and attempt to fill it out entirely.²

²Throughout the rest of this paper, the figures corresponding to the first model have names with the prefix `mod1`, and the figures corresponding to the second model have names with the prefix `mod2`.

Model 1

As specified earlier, for model 1 we assume we do not have any information about the lower part of the fast velocity layer, but that we know the geometry and magnitude of the slowness in the upper segment. Figure 6 depicts the starting (background) slowness model. The background wavefield is significantly distorted by the strong slowness variation in the background, as shown in Figure 7.

Figure 6: The starting slowness for model 1. This slowness model represents the background during the first pass (Figure 2). `paul3-mod1.Slow.1`
[CR]

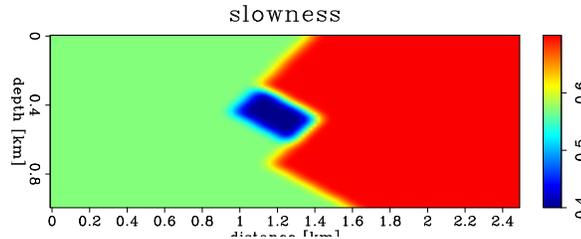
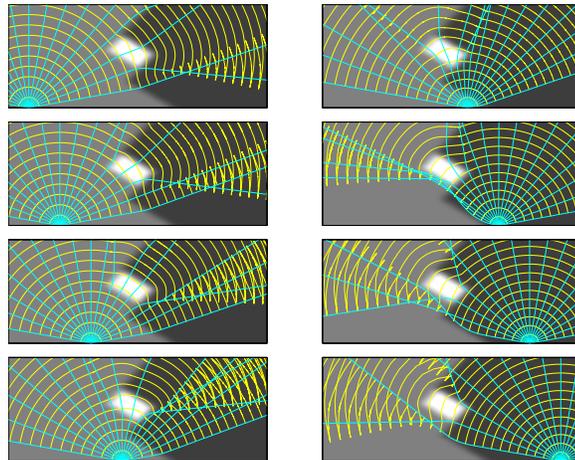


Figure 7: Rays and wavefronts for different source points in the subsurface. The wavefield triplicates because of the strong slowness variation in the background. `paul3-mod1.rays`
[CR]



Next, we migrate the synthetic data using the background slowness. The resulting image (Figure 8) is perfect in the upper part of the section, above the area with inaccurate slowness, but it is not completely accurate in the lower part, where the incorrect slowness model limits our imaging abilities. To analyze the quality of the image, we display it as angle-domain common image gathers (Prucha et al., 1999; Sava and Fomel, 2000) that have the property of showing flat events when the migration velocity is correct, and bending events when the velocity is incorrect.

In the next step, we compute an improved image (Figure 9) following the methodology outlined earlier in this paper and fully described in (Sava, 1999). As expected, there is no improvement in the upper section, since the velocity model is correct there, but only in the lower section, under the fault region. The picked velocity ratio surface is presented in Figure 10.

Finally, we can take the background (Figure 8) and enhanced (Figure 9) images and compute the image perturbation, depicted in Figure 11. For comparison, Figure 12 shows the theoretically-correct image perturbation computed using the true slowness to migrate the data.

Figure 8: The background image for model 1 presented as angle-domain common image gathers. The correctly migrated events are flat, while those incorrectly migrated are not, for example at the bottom of the lower layer. `paul3-mod1.bimg.1` [CR]

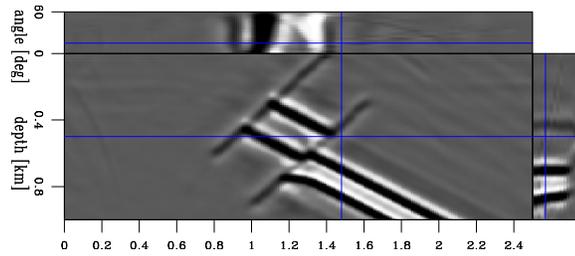
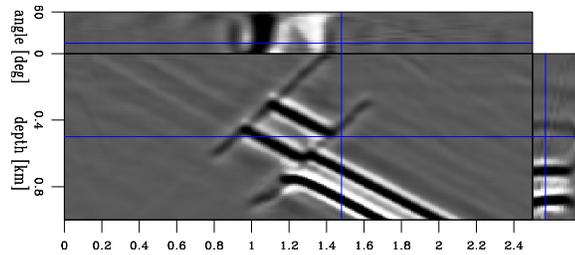


Figure 9: The enhanced image for model 1. `paul3-mod1.aimg.1` [CR]



Although the two images are not identical, they are similar. In practice, our goal is to come as close as possible to the image in Figure 12.

We have reached the point where we have created an image perturbation. We can now go one step further and compute the perturbation in slowness by solving the least-squares systems in Equations (1) or (2).

Figure 13 shows the result of the inversion process without any model regularization [$\epsilon = 0$ in equation (2)]. Every panel in the figure represents one iteration, as indicated by its title. The anomaly we seek is nicely recovered, with its position confined to the right area, as indicated by the schematic overlay. The absolute magnitude of the anomaly is smaller than that of the true one. This result can probably be explained by the nature of the Born approximation and the scaling of the velocity ratio surface. The definitive answer, however, is subject to further research.

Even though we have not yet imposed any smoothness constraint on the model, the anomaly is reasonably smooth, which is a direct consequence of the band-limited character of our wave-equation processing. The image is also free of major artifacts, although the bottom area of the model, which is not constrained by anything in the image perturbation, develops an anomaly of the opposite sign and another small, blob-like anomaly, possibly a wrap-around artifact.

Figure 10: The picked velocity ratio surface for model 1. This surface has values different from 1 only in the areas where the image has changed as a result of residual migration. `paul3-mod1.lmap.1` [CR]

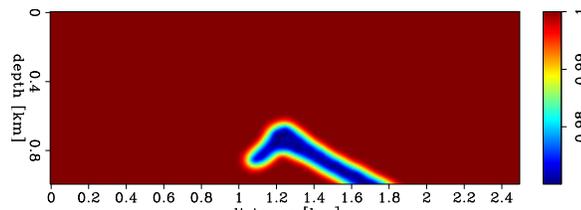


Figure 11: The image perturbation for model 1 computed from the background image and the image enhanced by residual migration. `paul3-mod1.dimg.1` [CR]

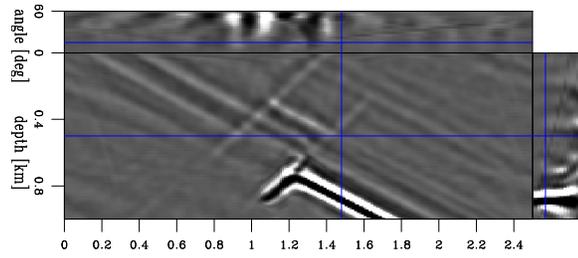


Figure 12: The perturbation in image for model 1 computed from the background image and the image obtained through migration with the true slowness. `paul3-mod1.dimg.0` [CR]

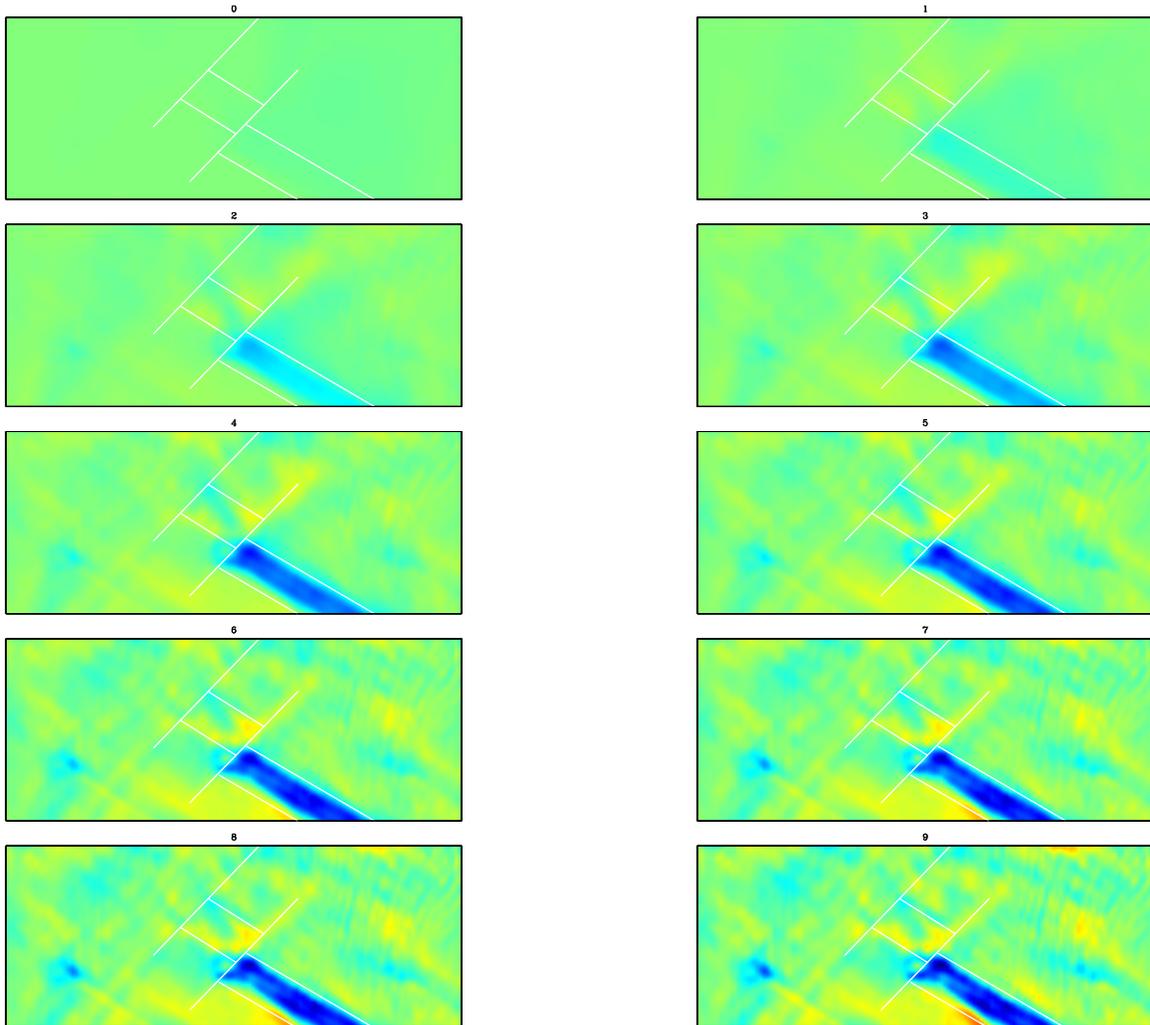
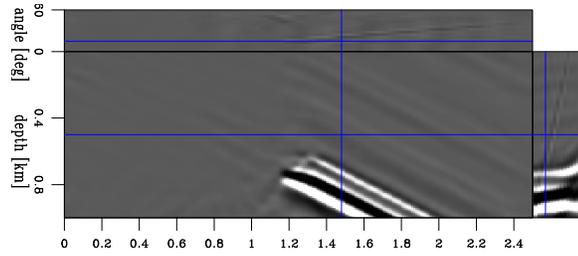


Figure 13: Iteration panels for the unregularized inversion in the case of model 1. `paul3-mod1.inv.plain.1` [CR]

Figure 14: Unregularized solution for model 1 after the first loop. `paul3-mod1.sol.plain.1` [CR]

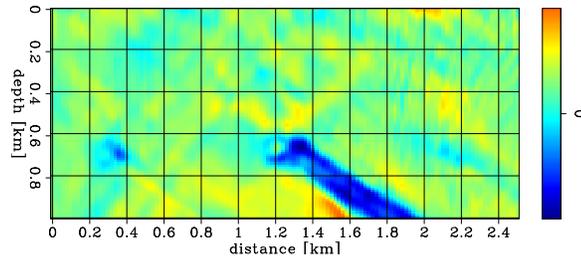


Figure 15 shows the result of model inversion for the regularized system (Equations 1). The anomaly is even smoother and confined to its correct location. Figures 14 and 16 show the final inversion result for the unregularized case and the case of preconditioning the model with a steering filter that forces the data to spread more along the geological dip and less perpendicular to it.

Finally, we update the background slowness model and repeat the loop. Figure 17 shows the updated background slowness, and Figure 18 the new background image. This new image should be compared to the previous background image (Figure 8). Notably, the bottom layer is pushed in depth closer to its correct position, and the angle-domain common image gathers show, on the average, flatter events, indicating that the new image is migrated with better velocity. The slowness perturbation is scaled differently before summation, as guided by the line-search optimization (Equation 3). We repeat the loop and obtain the slowness perturbation shown in Figure 19 and the updated image shown in Figure 20.

Here are the main conclusions we can derive from the first example:

- The image perturbation produced with residual migration is reasonably good, even for the case of a model with significant wavefield distortions in the background.
- As expected, wave-equation inversion produces smooth results, because of the intrinsic band-limited nature of wave propagation. Model regularization, however, helps to produce an even smoother slowness perturbation.
- The magnitude of the slowness anomaly we produced after inversion is less than the correct one. This is, in part, explained by the various approximations made when we created the image perturbation. It remains, nevertheless, an interesting topic that requires further investigation.
- The updated background slowness is able to produce an improved image, measured by the flatness of the angle-domain common image gathers. In other words, we have fulfilled the main goal of our procedure – image improvement – at the same time that we have improved the slowness model.
- Two WEMVA loops do not seem to be enough to recover the full slowness perturbation. We can get better results if we loop more, although cost remains a significant limitation. Here we have another direction of future work.

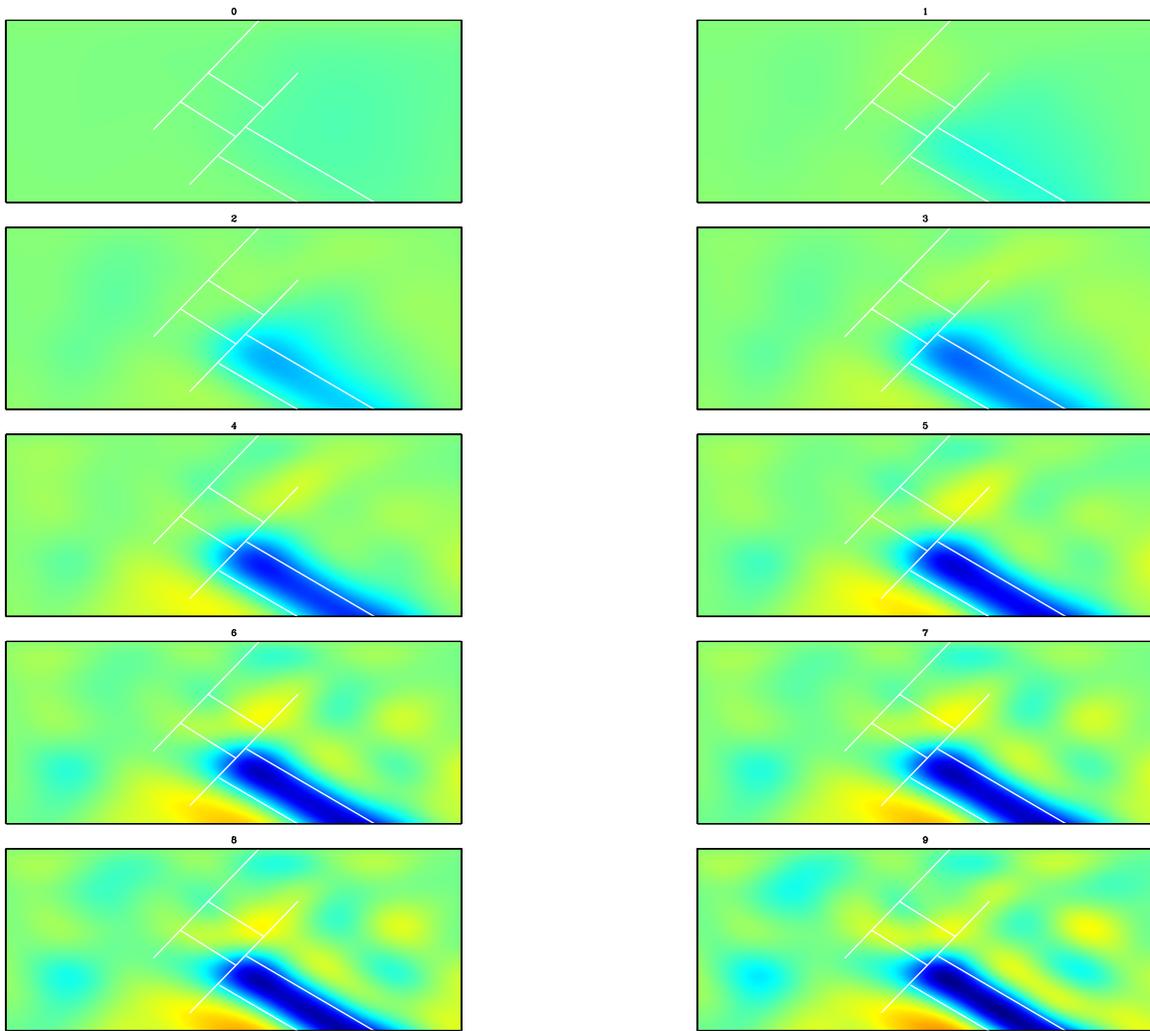


Figure 15: Iteration panels for the preconditioned inversion in the case of model 1. `paul3-mod1.inv.laplacian.1` [CR]

Figure 16: Solution for model 1 after the first loop in the case of regularization with steering filters. `paul3-mod1.sol.steering.1` [CR]

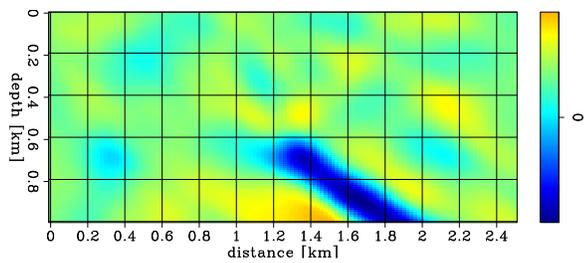


Figure 17: Updated background slowness for model 1 after the first loop. `paul3-mod1.Slow.2` [CR]

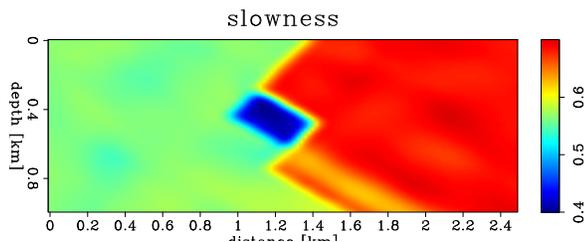


Figure 18: The updated background image for model 1 after the first loop. `paul3-mod1.bimg.2` [CR]

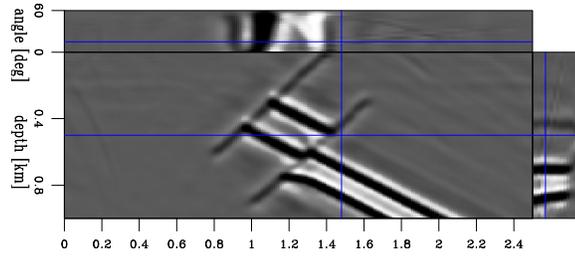
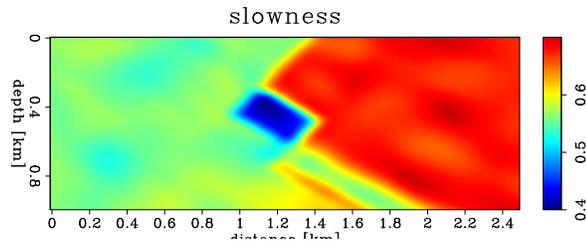


Figure 19: Updated background slowness for model 1 after the second loop. The line search procedure gives the slowness perturbation scaling parameter $\alpha = 0.625977$. `paul3-mod1.Slow.3` [CR]



Model 2

The second model we use to illustrate our procedure is derived from the same synthetic structure. This model differs from the first in that we now assume that we have no information about both the lower and the upper parts of the layer. The aim of this model is to recover the slowness anomaly, both in the upper and the lower sections.

For this second model, we follow the same basic steps as for the previous model. Here are the results:

1. Figure 21 represents the background slowness model. The fast layer is missing completely in this example.
2. Figure 22 represents rays and wavefronts for different shooting locations in the subsurface. The background wavefield is distorted, but not as much as in the case of the first model. This would indicate that this example is easier than the first one, although it probably doesn't make a significant difference because WEMVA handles naturally complex wavefields.
3. Figure 23 is the background image for model 2. Correctly migrated events are flat, while the incorrect ones are not. The background image is similar to the one in the

Figure 20: The updated background image for model 1 after the second loop. The line search procedure gives the slowness perturbation scaling parameter $\alpha = 1.0262$. `paul3-mod1.bimg.3` [CR]

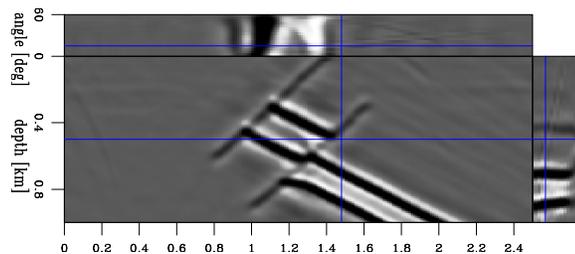


Figure 21: The starting slowness for model 2. It represents the background during the first pass (Figure 2). `paul3-mod2.Slow.1` [CR]

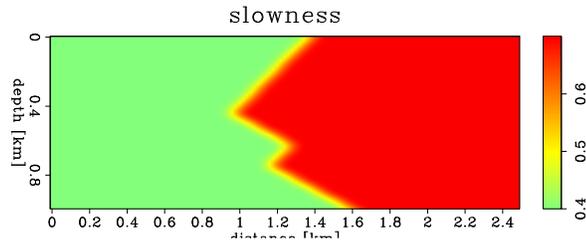
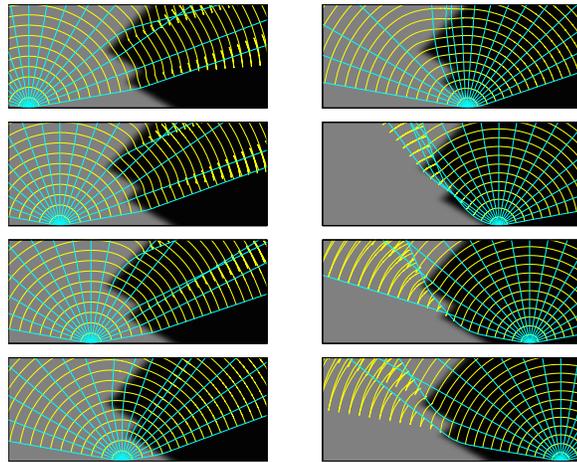
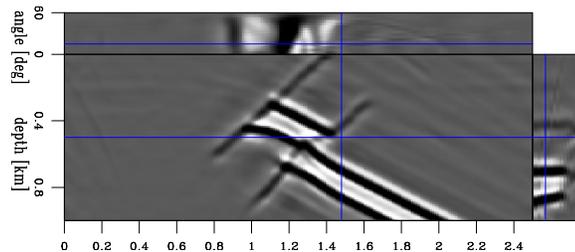


Figure 22: Rays and wavefronts for different source points in the subsurface. The wavefield for model 2 also triplicates because of the slowness variation in the background, but not as strongly as in the case of the first model. `paul3-mod2.rays` [CR]



preceding example (Figure 8), but it is distorted more, especially in the upper segment of the faulted layer.

Figure 23: The background image for model 2 presented as angle-domain common image gathers. Correctly migrated events are flat, while the incorrect ones are not. `paul3-mod2.bimg.1` [CR]



4. Figure 24 is the image enhanced by residual migration. Figure 25 is the corresponding picked velocity ratio surface.
5. Figure 26 is the image perturbation computed from the background image (Figure 23) and the image enhanced by residual migration (Figure 24). For comparison, Figure 27 is the image perturbation obtained through the difference of the images migrated with the true and background slowness functions.
6. Figure 28 are the iteration panels for unregularized inversion, and Figure 29 are the iteration panels for preconditioned inversion. Here we can see the major differences and similarities between this and the first example: the lower part of the layer is well recovered; the upper part of the layer is also recovered reasonably well, although not as well as the lower part. The region around the fault, however, is not constrained enough

Figure 24: The enhanced image for model 2. `paul3-mod2.aimg.1` [CR]

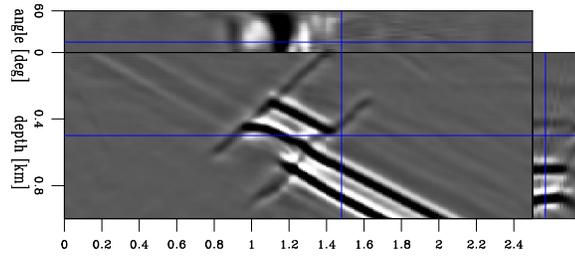


Figure 25: The picked velocity ratio surface for model 2. `paul3-mod2.lmap.1` [CR]

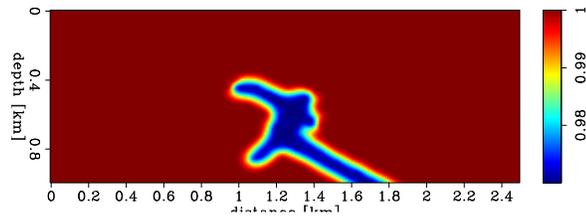


Figure 26: The image perturbation for model 2 computed from the background image and the image enhanced by residual migration. `paul3-mod2.dimg.1` [CR]

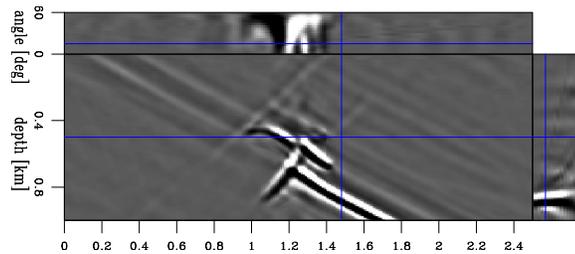
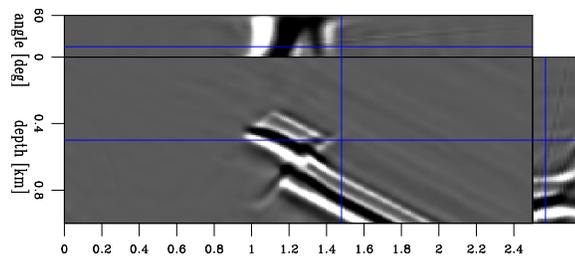


Figure 27: The image perturbation for model 2 computed from the background image and the image obtained through migration with the true slowness. `paul3-mod2.dimg.0` [CR]



and shows high spatial-frequency variation in slowness (Figure 28). However, if we regularize the model during the inversion, we recover a much smoother anomaly. The most significant change occurs in the area surrounding the fault. In this case, we do not observe the large fluctuations that occur in the unregularized example, although the regularization operator causes the slowness anomaly to leak outside its bounds.

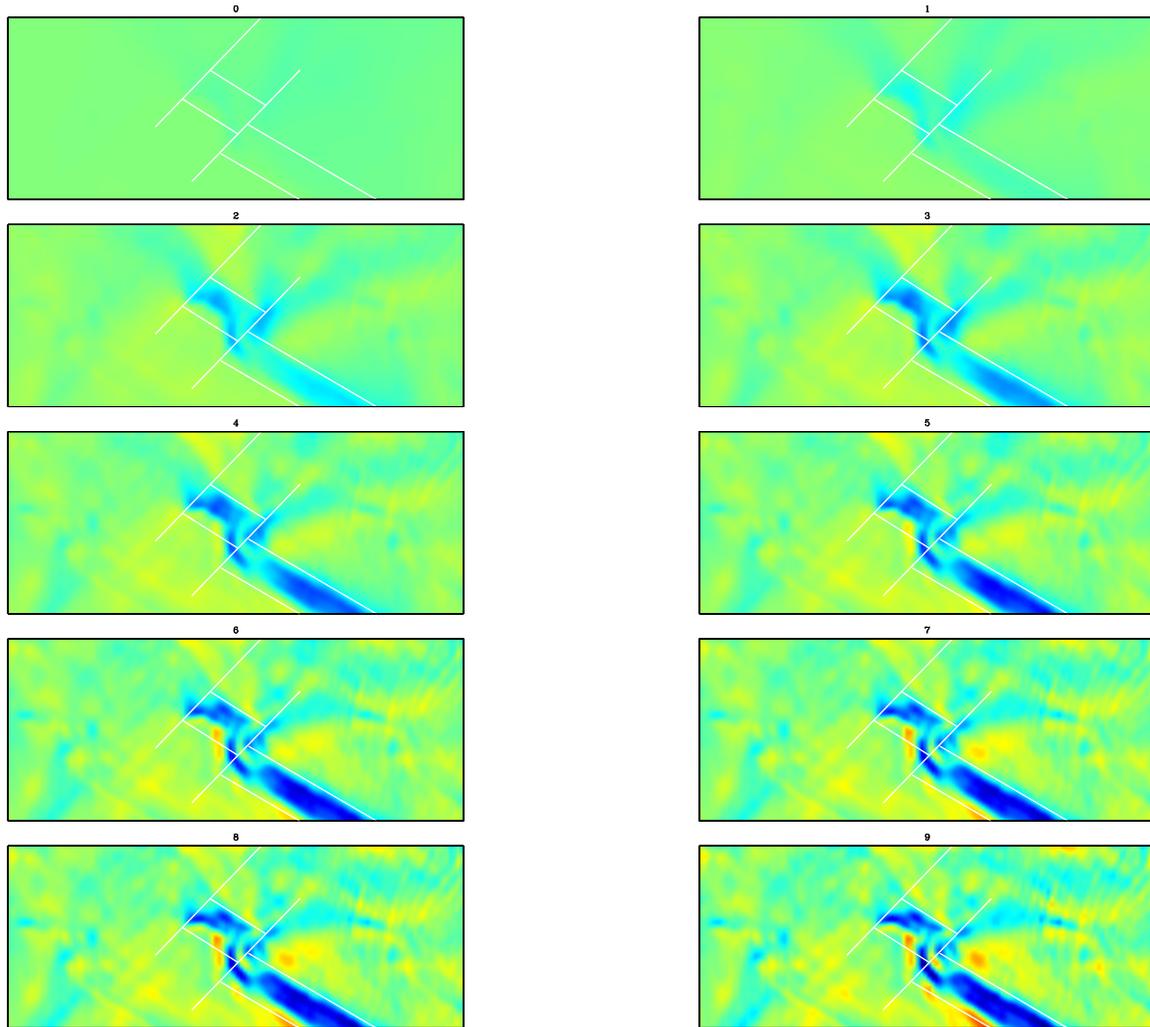


Figure 28: Iteration panels for the unregularized inversion in the case of model 2. `paul3-mod2.inv.plain.1` [CR,M]

7. Figure 30 is the unregularized solution after the first loop, and Figure 31 is the solution after the first loop when the model is regularized using steering filters. The second of the two is a significantly improved result: the anomaly is much smoother, and the leakage in the vertical direction outside the layer bounds is reduced compared to the case of preconditioning with an isotropic Laplacian operator.
8. Figure 32 is the updated background slowness after the first loop, and Figure 33 is the corresponding updated background image. Although not yet completely there, the fast

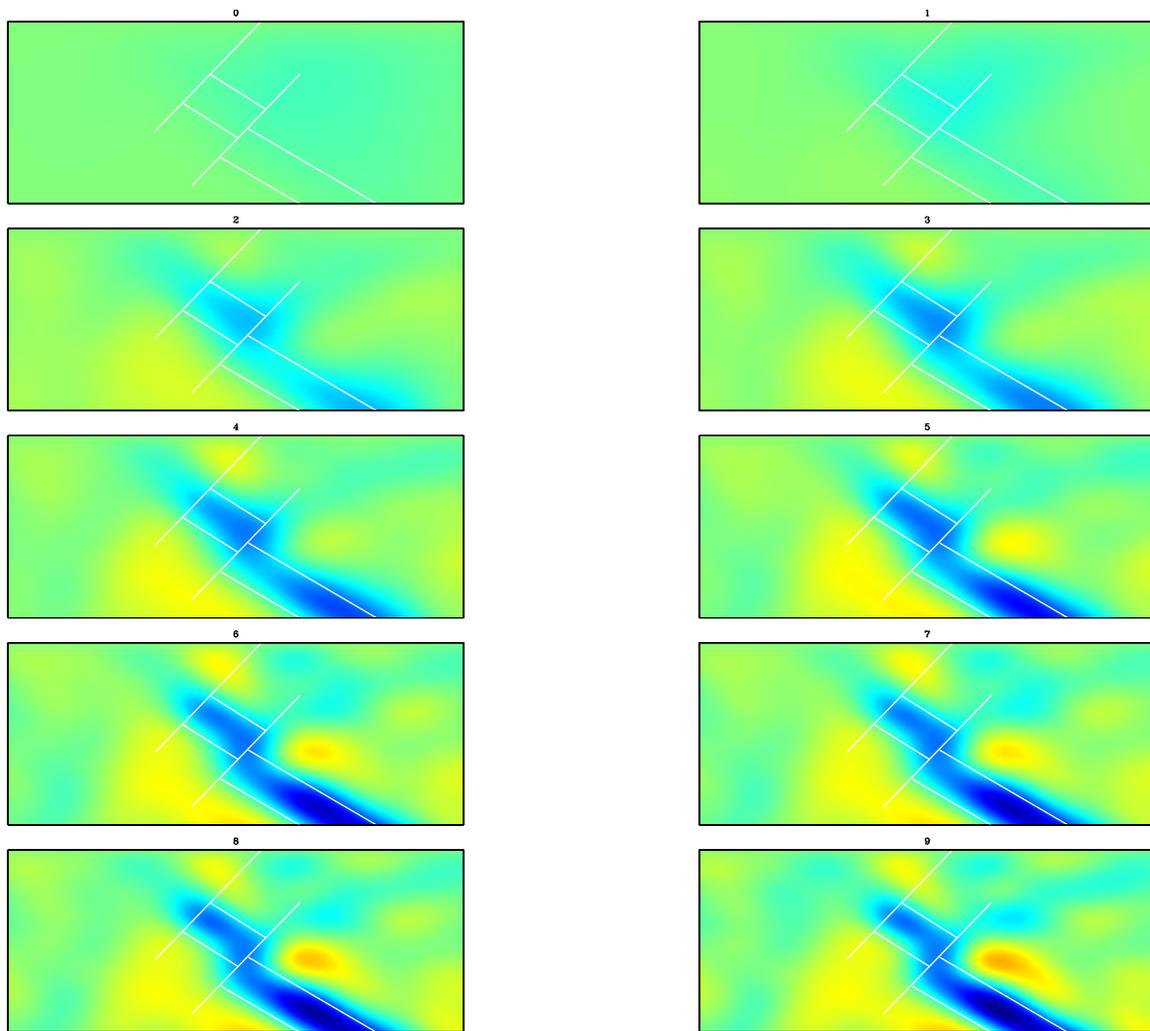


Figure 29: Iteration panels for the preconditioned inversion in the case of model 2.
`paul3-mod2.inv.laplacian.1` [CR,M]

Figure 30: Unregularized solution for model 2 after the first loop.
`paul3-mod2.sol.plain.1` [CR]

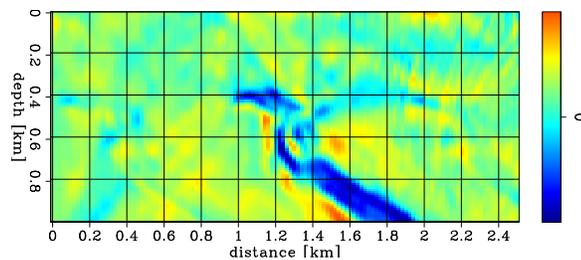
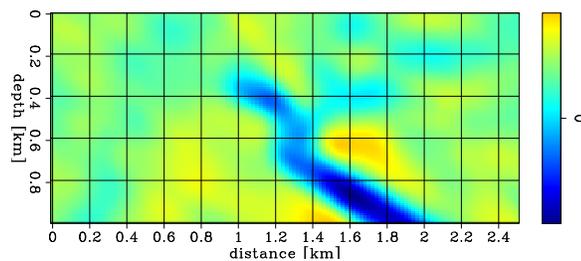


Figure 31: Solution for model 2 after the first loop in the case of regularization with steering filters.
`paul3-mod2.sol.steering.1` [CR]



layer begins to build-up, and, consequently, the quality of the migrated image improves. However, we have not reached a satisfactory image at every location; therefore, we need to repeat the WEMVA loop. When we repeat the loop we obtain the slowness perturbation shown in Figure 34 and the updated image shown in Figure 35.

Figure 32: Updated background slowness for model 2 after the first loop. The line search procedure gives the slowness perturbation scaling parameter $\alpha = 1.64043$.
`paul3-mod2.Slow.2` [CR]

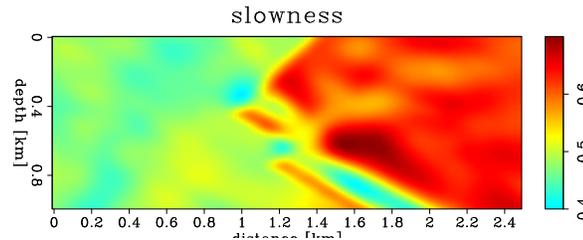


Figure 33: The updated background image for model 2 after the first loop.
`paul3-mod2.bimg.2` [CR]

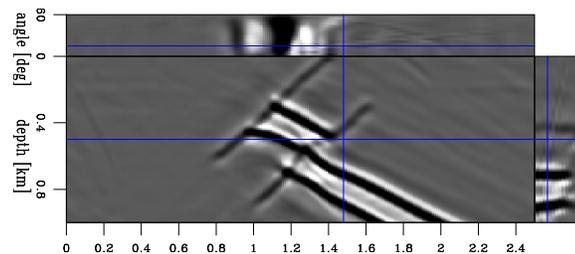
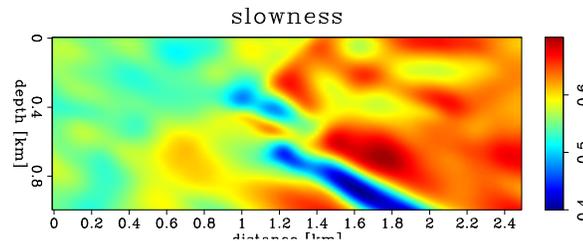


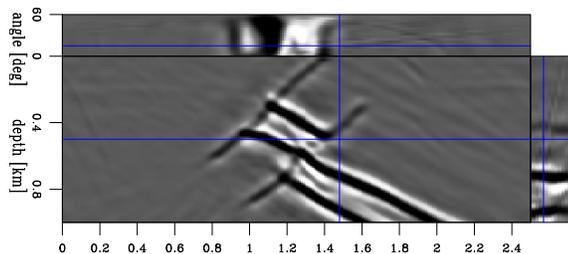
Figure 34: Updated background slowness for model 2 after the second loop. The line search procedure gives the slowness perturbation scaling parameter $\alpha = 2.81967$.
`paul3-mod2.Slow.3` [CR]



The main conclusions we can derive from the second example are:

- Residual migration is successful in producing a good image perturbation, even though the starting image is more complex and significantly more distorted than in the first case.
- The anomaly produced by WEMVA without model regularization does not have as high a quality as it does in the original case. Large distortions in the original image, possibly coupled with residual migration distortions, lead to high spatial frequency variations of the inverted anomaly. We can, however, control the shape of the anomaly with smoothness constraints on the model, among which steering filters produce a promising result.
- The updated image is characterized by flatter angle-domain common image gathers. One loop, however, doesn't seem to be enough to achieve a completely satisfactory result, and we need to iterate more.

Figure 35: The updated background image for model 2 after the second loop. `paul3-mod2.bimg.3` [CR]



CONCLUSIONS

In this paper, we demonstrated the feasibility of migration velocity analysis using wave-equation methods. The results we obtained on 2-D synthetic models are very encouraging. Cost remains a major limitation, although as our method progressively matures, we have more and more computer power at our disposal.

The main point we would like to make is that unlike other velocity inversion techniques, our method updates the velocity model by improving the quality of the migrated image, and not by fitting the recorded data. It therefore takes full advantage of the intertwined nature of migration and velocity analysis,

Wave-equation migration techniques are known for their potential to handle complex wavefields. Since our velocity analysis method is also based on processing with the wave-equation, it inherits stability and constrains the derived velocity to be smoother than what a travelttime-based method would allow. Furthermore, we can control the shape of our derived velocity anomalies by imposing external constraints, either model-independent, like Laplacian smoothing, or model-dependent steering filters.

The Born approximation on which we base our method limits the amount of improvement we can allow on the starting image. For lack of a better procedure, we now chose to continue with a technique in which we scale-down the image perturbation and later scale-up the slowness perturbation. Handling the limitations imposed by the Born approximation is one of the most useful and exciting areas of future research.

In addition to improvements in scaling and rescaling the image and slowness perturbations, other potential directions of future research include semblance and differential semblance analysis for more reliable image enhancements and an analysis of the implications of the complex nature of the wavefields, specifically the benefits of implicitly using all the arrivals in the wavefield for velocity analysis rather than merely a single one.

REFERENCES

- Biondi, B., and Sava, P., 1999, Wave-equation migration velocity analysis: SEP-100, 11–34.
- Claerbout, J., 1999, Geophysical estimation by example: Environmental soundings image enhancement: <http://sepwww.stanford.edu/sep/prof/>.

Clapp, R. G., and Biondi, B. L., 1998, Regularizing time tomography with steering filters: SEP-97, 137-146.

Gray, S. H., 1998, Speed and accuracy of seismic migration methods: Mathematical Geophysical Summer School, Stanford University.

Prucha, M. L., Biondi, B. L., and Symes, W. W., 1999, Angle-domain common image gathers by wave-equation migration: SEP-100, 101-112.

Sava, P., and Fomel, S., 2000, Angle-gathers by Fourier Transform: SEP-103, 119-130.

Sava, P., 1999, Enhancing common-image gathers with prestack Stolt residual migration: SEP-102, 47-60.

APPENDIX

This appendix is a picture summary of the processing we describe in the paper. We present four figures for each of the two models we discuss in the earlier sections.

- Figures (A-1,A-2) and (A-5,A-6) describe the first two WEMVA loops (Figure 2) for models 1 and 2, respectively. From top to bottom, the panels depict
 - the background slowness model,
 - the background image,
 - the image improved by residual migration, and
 - the perturbation in image derived from the background and improved images.
- Figures A-3 and A-7 describe the changes on the slowness model for the first two WEMVA loops (Figure 2). From top to bottom, the panels depict
 - the correct slowness,
 - the background slowness,
 - the updated slowness after the first loop, and
 - the updated slowness after the second loop.
- Figures A-4 and A-8 describe the changes that are induced on the image for the first two WEMVA loops (Figure 2). From top to bottom, the panels depict
 - the image migrated with the true slowness,
 - the image migrated with the starting slowness,
 - the image migrated with the slowness updated after the first loop, and
 - the image migrated with the slowness updated after the second loop.

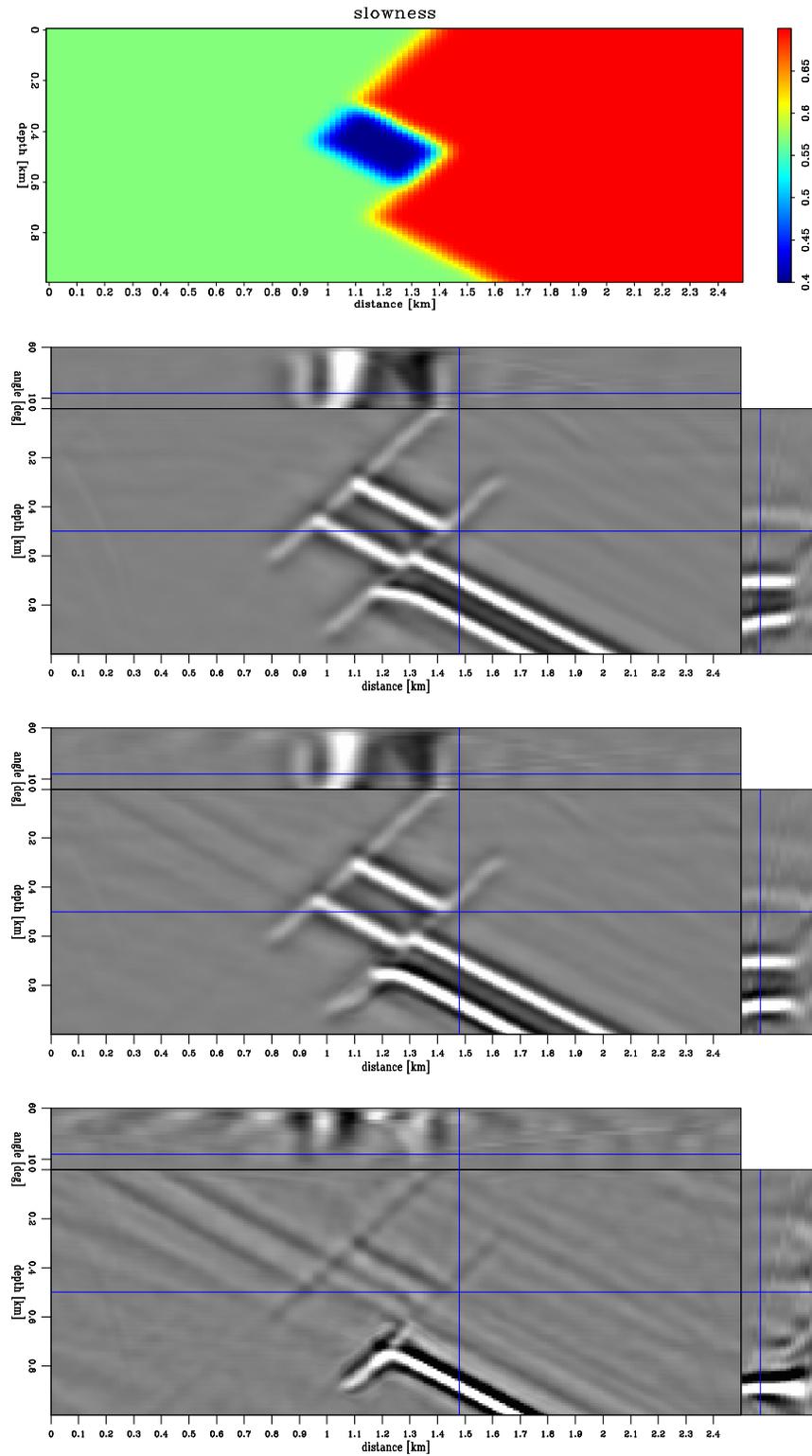


Figure A-1: Model 1, WEMVA loop 1. From top to bottom: background slowness, background image, improved image, and image perturbation. `paul3-mod1.loop.1` [CR]

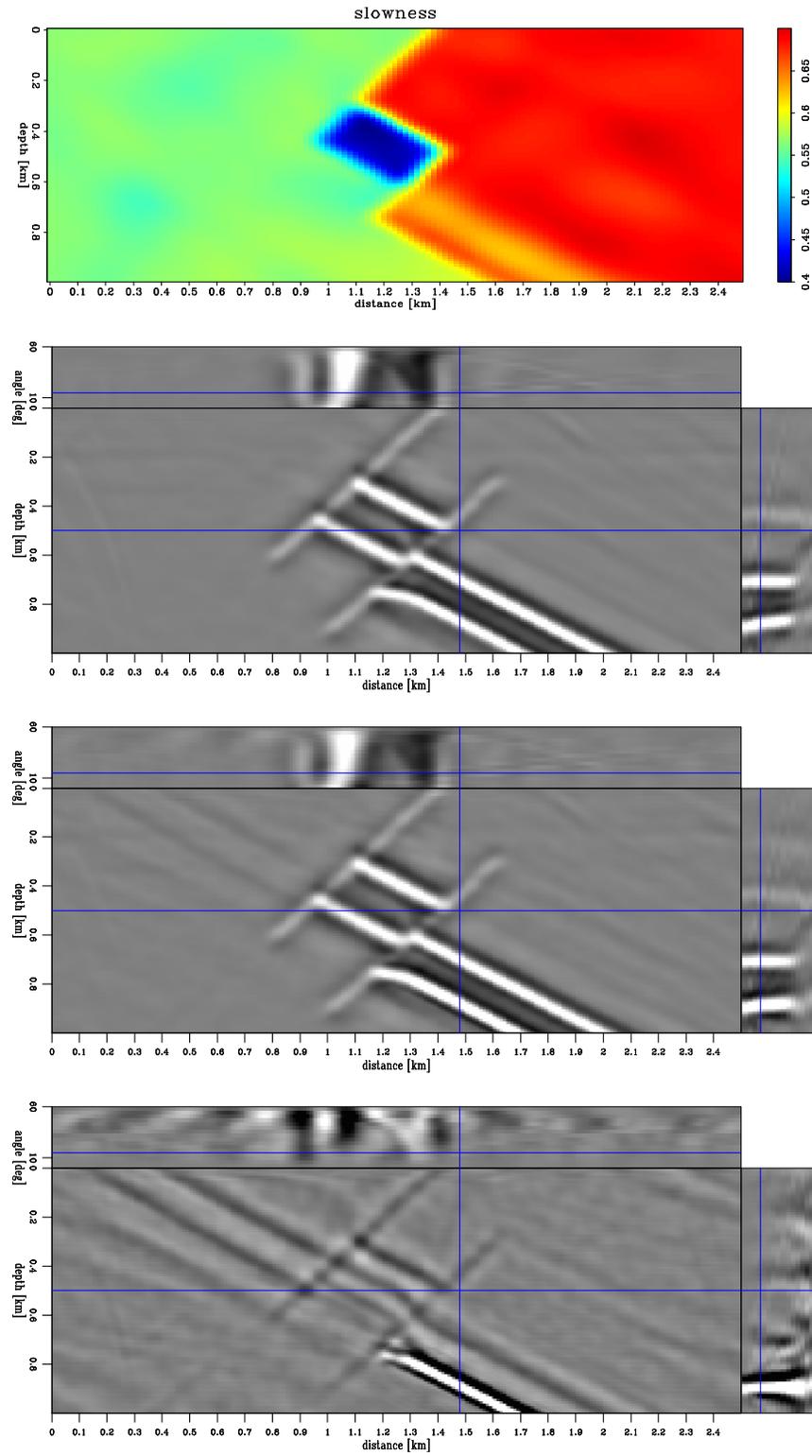


Figure A-2: Model 1, WEMVA loop 2. From top to bottom: background slowness, background image, improved image, and image perturbation. `paul3-mod1.loop.2` [CR]

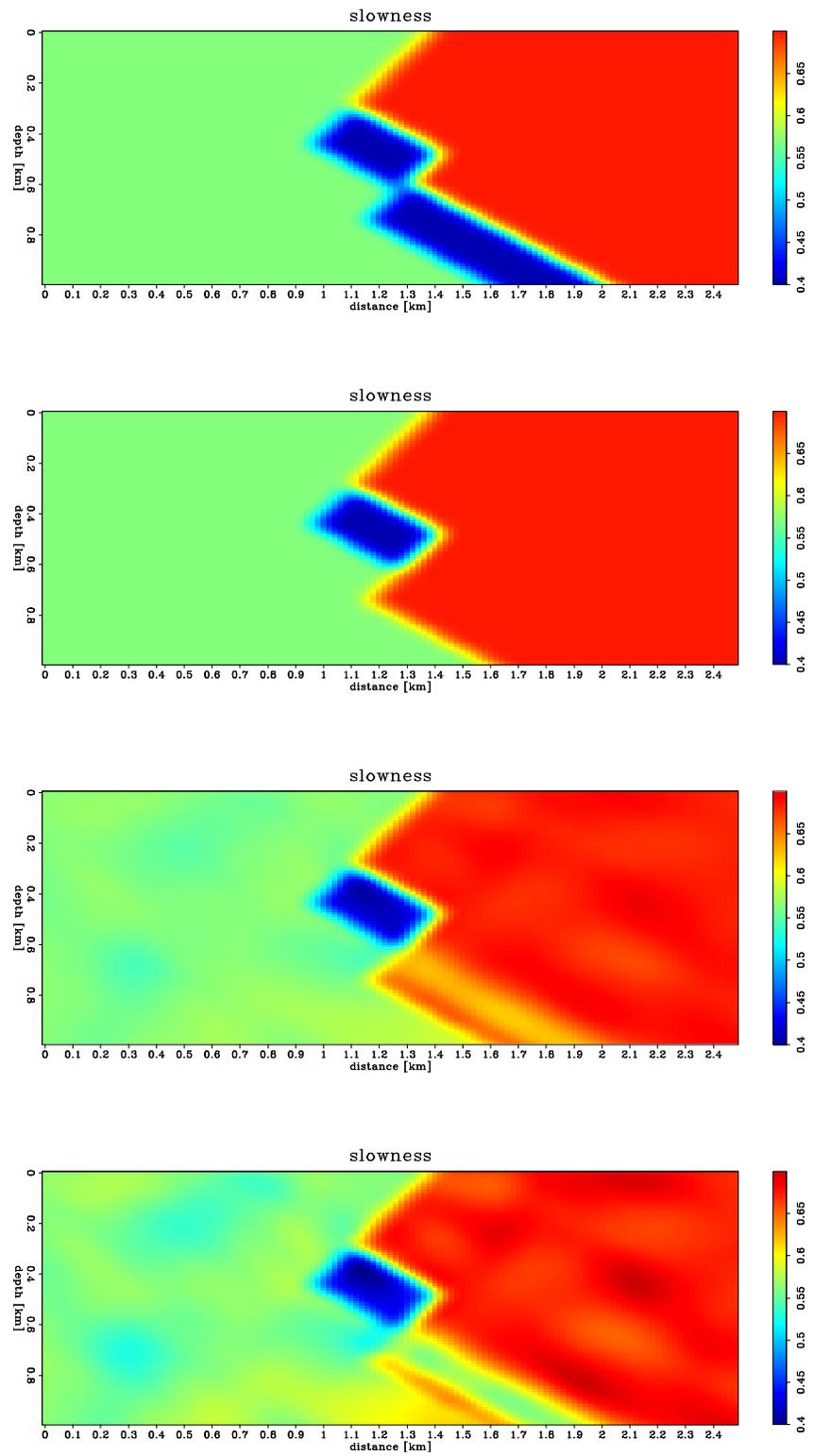


Figure A-3: Model 1, summary of slowness changes with loop number. From top to bottom: the true slowness, the starting slowness, the slowness updated after the first loop, the slowness updated after the second loop. `paul3-mod1.slows` [CR]

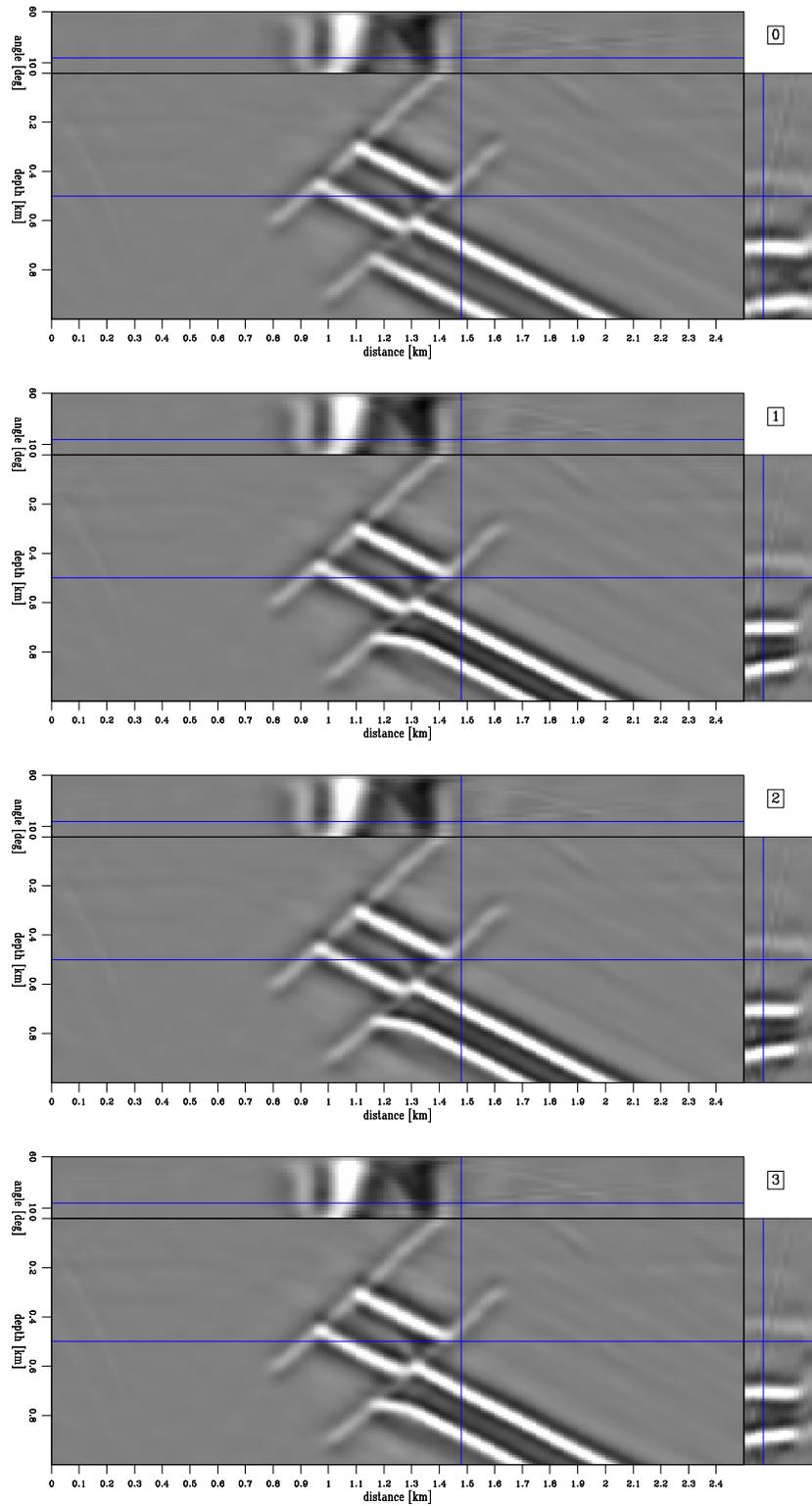


Figure A-4: Model 1, summary of image changes with loop number. From top to bottom, images obtained with the true slowness, the starting slowness, the slowness updated after the first loop, the slowness updated after the second loop. `paul3-mod1.iter` [CR]

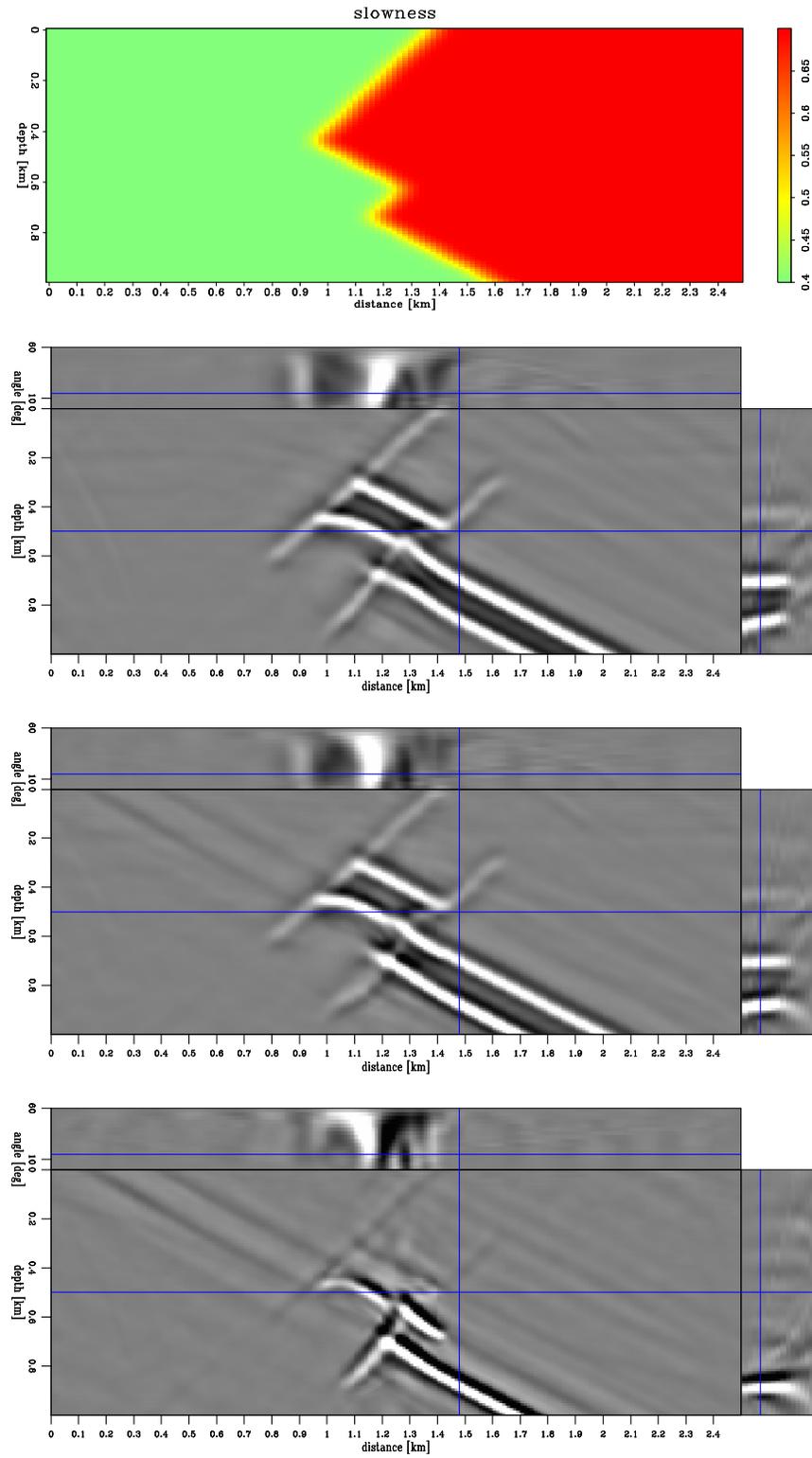


Figure A-5: Model 2, WEMVA loop 1. From top to bottom: background slowness, background image, improved image, and image perturbation. `paul3-mod2.loop.1` [CR]

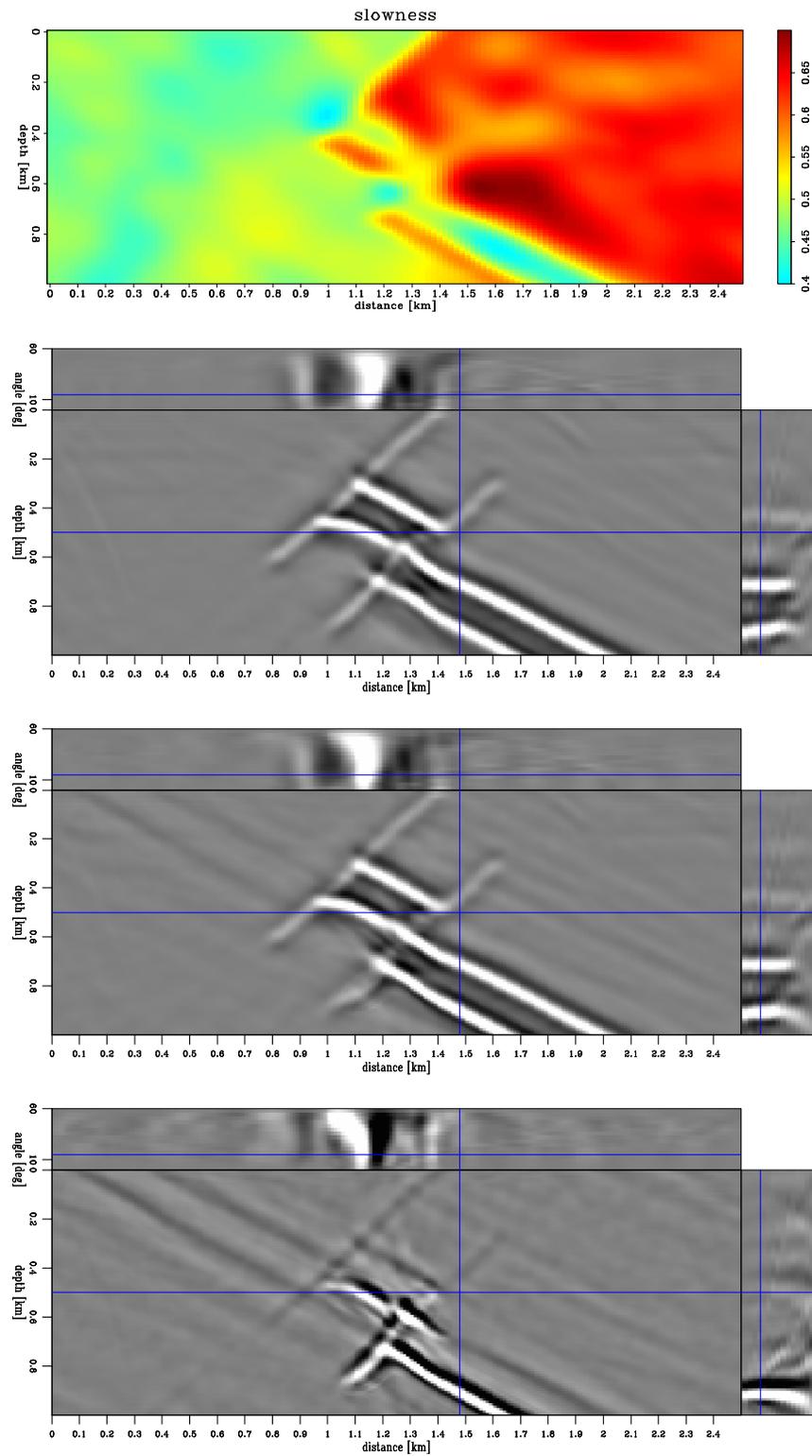


Figure A-6: Model 2, WEMVA loop 2. From top to bottom: background slowness, background image, improved image, and image perturbation. `paul3-mod2.loop.2` [CR]

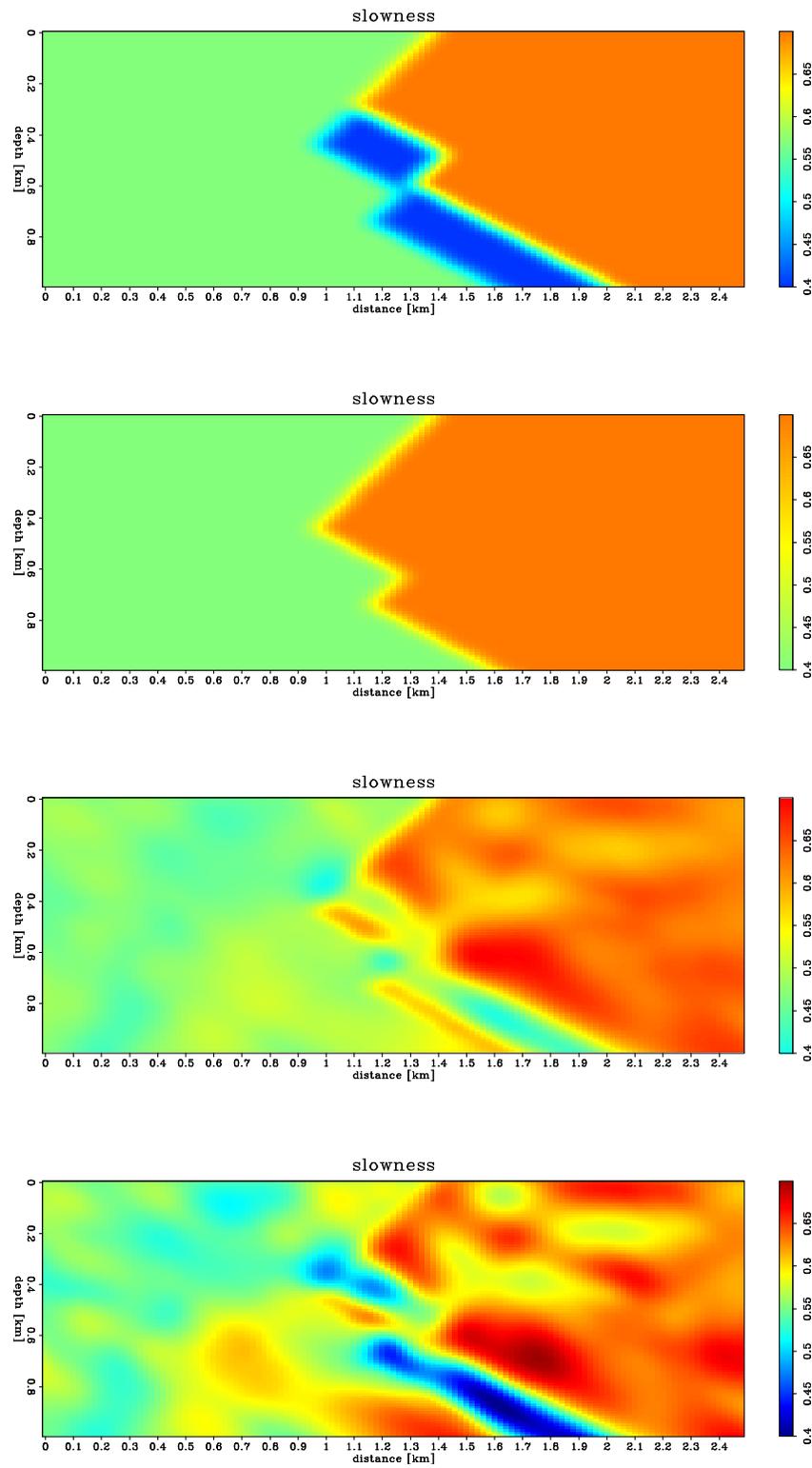


Figure A-7: Model 2, summary of slowness changes with loop number. From top to bottom: the true slowness, the starting slowness, the slowness updated after the first loop, the slowness updated after the second loop. `paul3-mod2.slows` [CR]

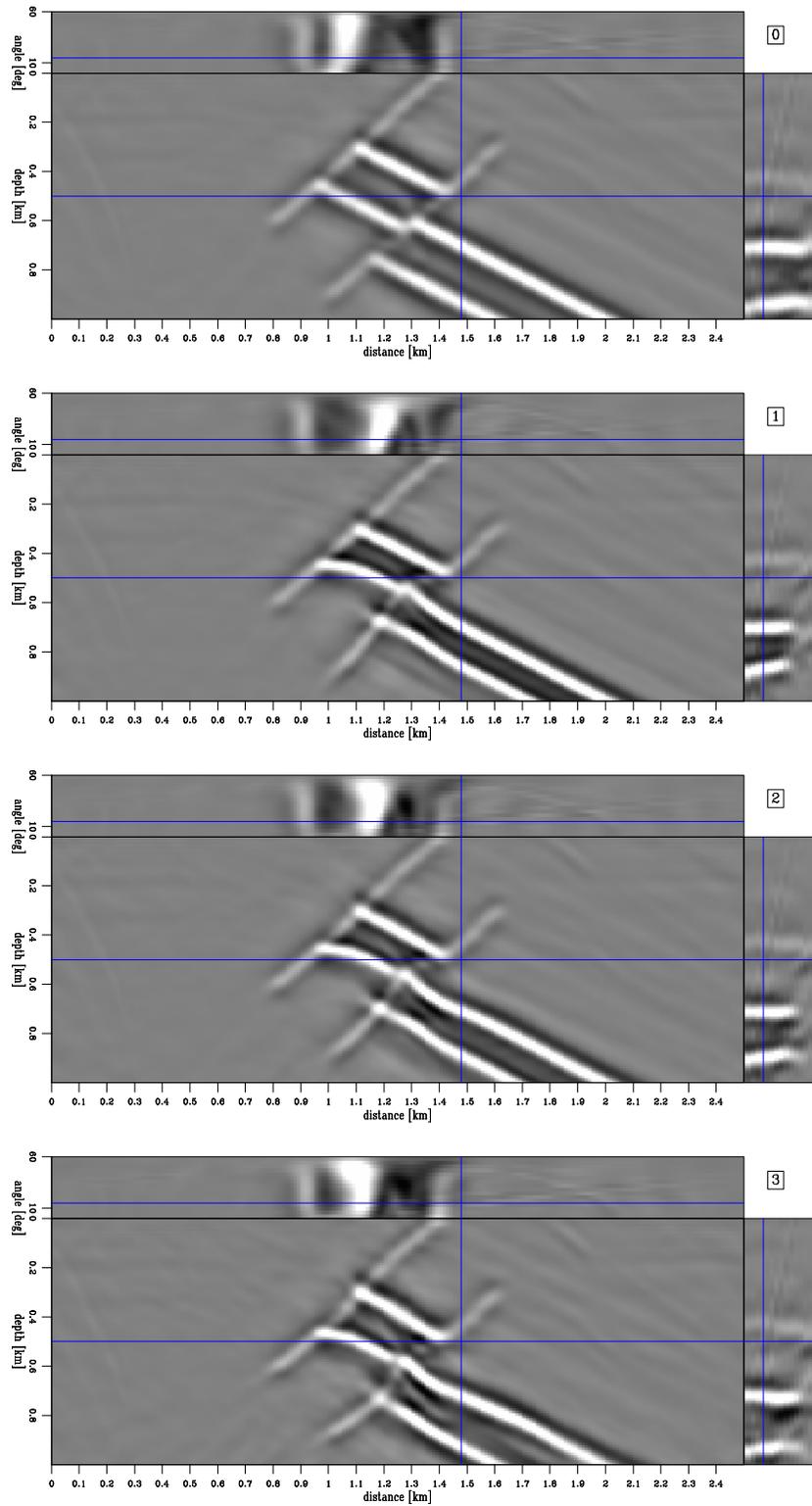


Figure A-8: Model 2, summary of image changes with loop number. From top to bottom, images obtained with the true slowness, the starting slowness, the slowness updated after the first loop, the slowness updated after the second loop. `paul3-mod2.iter` [CR,M]

Regularizing tomography with non-stationary filters

Robert G. Clapp¹

ABSTRACT

The ideal regularizer is the inverse of the model covariance matrix. Often the model covariance matrix has a complicated structure that is difficult to characterize. Non-stationary prediction error filters (PEF) have the ability to describe complicated model behavior. Non-stationary filters are effective regularizers for missing data and tomography problems.

INTRODUCTION

Most geophysical problems are either under-determined or mixed-determined, requiring some type of regularization. The ideal regularizer is the inverse model covariance (Tarantola, 1986). In previous papers I have shown that a space-varying operator composed of small plane-wave annihilation filters, or *steering filters*, can be an effective regularization operator (Clapp et al., 1997, 1998; Clapp and Biondi, 1998, 2000). Steering filters are best suited to describing models with relatively simple covariance functions. For a certain class of velocity models, such as models with discontinuities, steering filters have difficulty accurately describing model covariance.

PEFs are able to describe a much wider class of models than steering filters. To robustly estimate a PEF we must have a model with stationary statistics, something that is rarely true with seismic problems. We can often satisfy the stationarity requirement by breaking up our problem into small patches (Claerbout, 1992b). Unfortunately, we can only make our patch size so small before we can't generate sufficient statistics to estimate our PEF (Crawley et al., 1999).

An alternative approach, proven to be effective when dips change quickly, is to estimate PEFs in micro patches with a non-stationary PEF (Crawley et al., 1999; Clapp and Brown, 1999, 2000). When dealing with discontinuities, regularizing with a non-stationary PEF can be more effective in describing the model covariance than steering filters. Non-stationary filters do a better job honoring sharp boundaries and characterizing complex models.

I will begin by showing how steering filters perform poorly at discontinuities. I will then show how to build and estimate a non-stationary PEF. I will use the non-stationary PEF in the context of a missing data problem. I will conclude by using a non-stationary PEF for

¹email: bob@sep.stanford.edu

regularization to tomography.

BACKGROUND

In general, geophysical inverse problems (inverting for some model (\mathbf{m}) , given some (\mathbf{d}) , while applying some operator (\mathbf{L})) are ill-posed. A classic example of this is the missing data problem (Claerbout, 1999). The goal of the missing data problem is to interpolate intelligently between a sparse set of known points. For example, let's take a synthetic velocity model with an upper horizontal reflector, an anticline between two unconformities, and updipping layer at the bottom of the model. Suppose we have velocity measurements at several wells (Figure 1) and you would like to interpolate it onto a regular 2-D mesh.

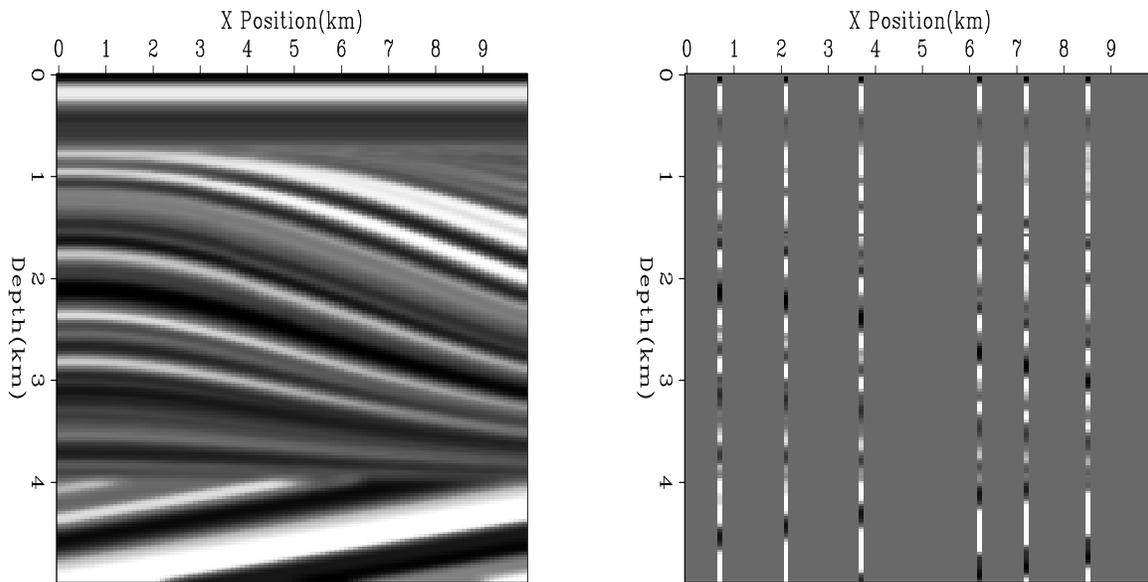


Figure 1: Left panel shows a synthetic velocity model, right panel shows a subset of that data chosen to simulate well log data. `bob3-well-logs` [ER]

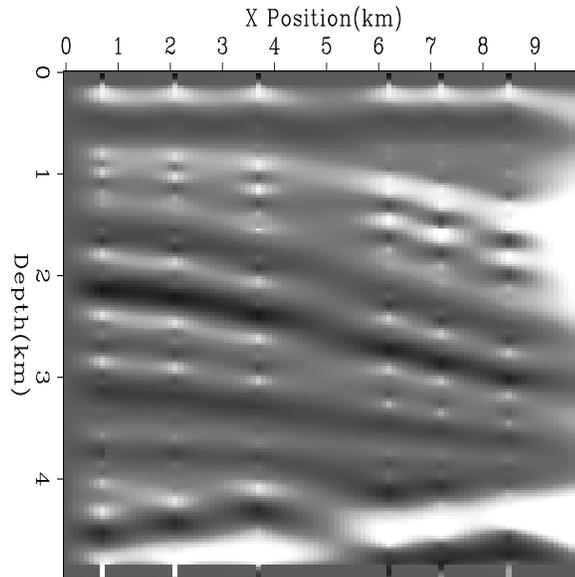
The geophysicist might follow the approach described by Claerbout (1999), first interpolating the irregular data onto a regular mesh by applying some type of binning operator, \mathbf{B} , then defining a fitting goal that requires the model to fit the data exactly at the known points (\mathbf{J}) ,

$$\mathbf{J}\mathbf{B}\mathbf{d} \approx \mathbf{J}\mathbf{m}. \quad (1)$$

At model locations where there are no data values, we want the model to be ‘smooth’, therefore we will use Tikhonov regularization to minimize the output of a roughening operator applied to the model,

$$\mathbf{0} \approx \mathbf{A}\mathbf{m}. \quad (2)$$

Figure 2: Interpolation result after 200 iterations using an inverse Laplacian regularization operator. Note the edges effects at the top and bottom of the model due to using a internal convolution operator. `bob3-qdome-lap` [ER]



If we don't have any other knowledge about our model, an isotropic operator like the Laplacian might be a logical choice for \mathbf{A} since it leads to the “minimum energy” solution. If I apply the fitting goals implied by (1) and (2) for 200 iterations using the Laplacian for \mathbf{A} I get Figure 2. The result is what has been euphemistically referred to as the ‘ice cream cone result’ (Brown, 1998). By spreading information isotropically, the model goes smoothly from our known points to some local average. We see little to no continuation of layers, which is generally a thoroughly unsatisfactory result.

Covariance

With no other information, the Laplacian might be the best regularization operator that we could use. But if we know something else about our model, can we do better? According to Tarantola (1987), we should be using the inverse model covariance for our regularization operator. The statistics of the model vary spatially, but by breaking up into four patches, one above the anticline, one below the anticline, and two within the anticline we can at least approximate stationary statistics. If we calculate the covariance within patches where the statistics are relatively stable, we get Figure 3.

Steering filters

If we examine our desired model, it is apparent that the covariance function varies within at least two of our four patches (we can also see this in the covariance function of patch 2 and 4 of Figure 3). Therefore, it follows that we should get a better image by making smaller and smaller patches. Crawley (1998) showed that this is true when solving a data interpolation problem. Traditional methods for characterizing the model like PEFs and variograms can only grow so small before we have insufficient statistics to calculate them.

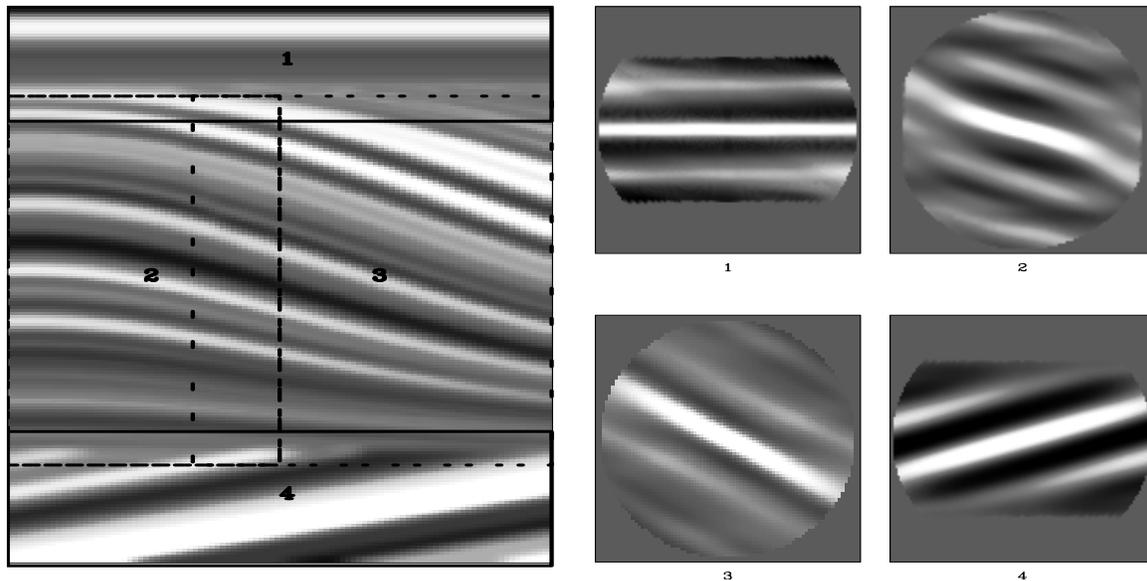


Figure 3: The covariance at four different regions of our model (left panel of Figure 1.) The top left is above the upper unconformity; top right, the upper portion of the anticline; bottom left, lower portion of the anticline; and bottom right, below the lower anticline. `bob3-covar-change` [ER]

When our stationary assumption is valid, such as in regions one and four of Figure 3 the covariance matrix is fairly simple. We have a primary trend oriented along the dip of the velocity field that slowly dies out and a ringing effect due to the sinusoidal nature of our model. We would like to come up with a way to emulate the primary trend of the covariance matrix through minimal information.

To do this it is important to remember that our regularization operator should have the inverse spectrum of the covariance matrix. Therefore if the covariance function is primarily a dipping event, our regularization operator should be destroying that dip. Claerbout (1990; 1992a) showed how to estimate the primary dip in a region and how to construct a filter that could destroy that dip. These small filters, which I refer to as *steering filters* can be as simple as a two or three point filter, Figure 4. A steering filter consists of a fixed '1' and one or more coefficients in the next column. The location of the filter coefficients in the second column determines the dip that the filter will destroy. Figure 5 is the inverse impulse response of Figure 4. Note how the general orientation of the impulse response is approximately the same as the covariance function below the lower unconformity. If we assume that velocity

Figure 4: A steering filter which annihilates dips of 22.5 degrees. `bob3-small-filter` [NR]

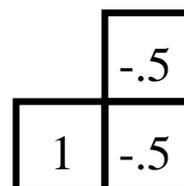
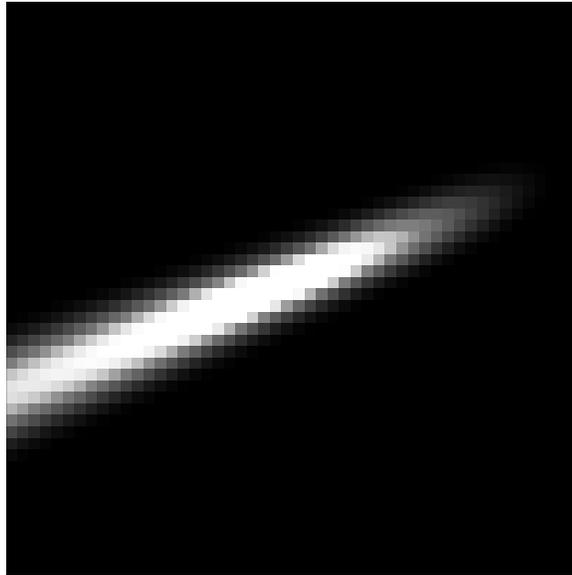


Figure 5: The impulse response of the inverse of Figure 4.
bob3-small-response [ER]



follows structure and we have some guesses at reflector position, we can use this information to build our steering filters. For this problem we will assume that we have the location of four reflectors, one above the top unconformity, two between the unconformities and one below the lower unconformity (left-panel of Figure 6). If we interpolate these dips to our entire model space we have all we need to construct a space-varying operator composed of steering filters. If we use this operator as our regularizer, we get Figure 7 as our interpolation result. The steering filters did a significantly better job than the isotropic regularizer.

DISCONTINUITIES AND STEERING FILTERS

Now let's move onto a model with discontinuities. Figure 8 is similar to Figure 1, with the exception that we now have a listric fault in the middle of the anticline structure. If we follow the same interpolation path as we did in the last section, the right panel of Figure 8 is our interpolation result. Instead of reproducing the fault we have created a model that smoothly changes from horizons on the left of the fault to the corresponding horizons to the right side of the fault. In many cases a smooth change is not only acceptable, but desirable (for example, ray-based methods require a smooth model). In other cases the smooth change is unrealistic and something we want to avoid (salt boundaries and some fault boundaries). Our interpolation fails at the fault because we are not correctly describing the covariance along the edge of the fault. The problem is that the covariance function at the edge of the fault is not symmetric. Along the left edge of the fault we have good correlation with points to the left but our correlation with points to the right our correlation is shifted. This asymmetric behavior is difficult to describe with steering filters.

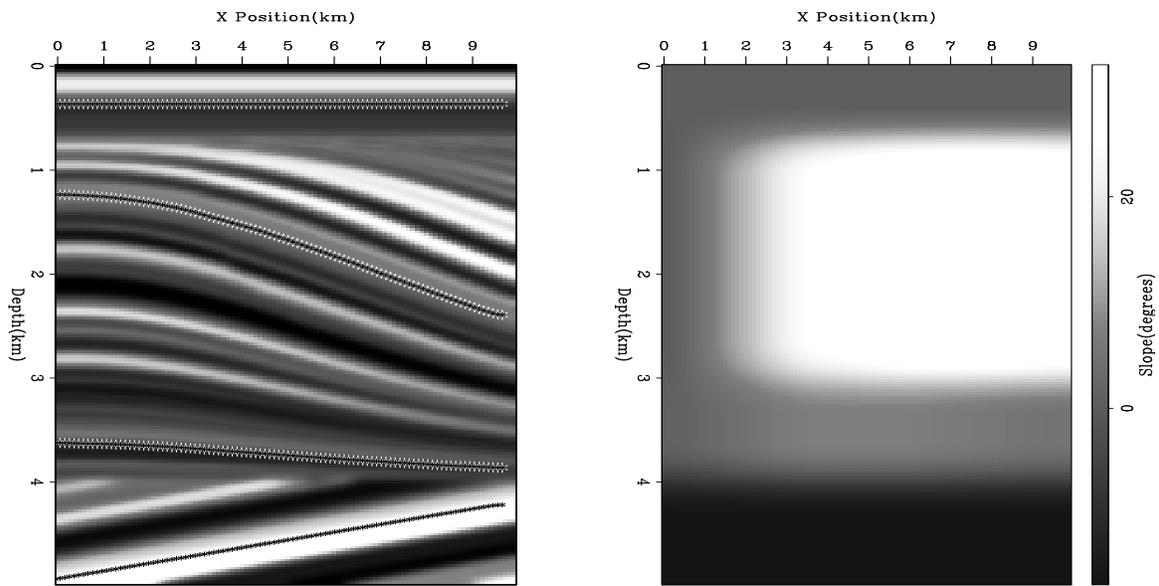
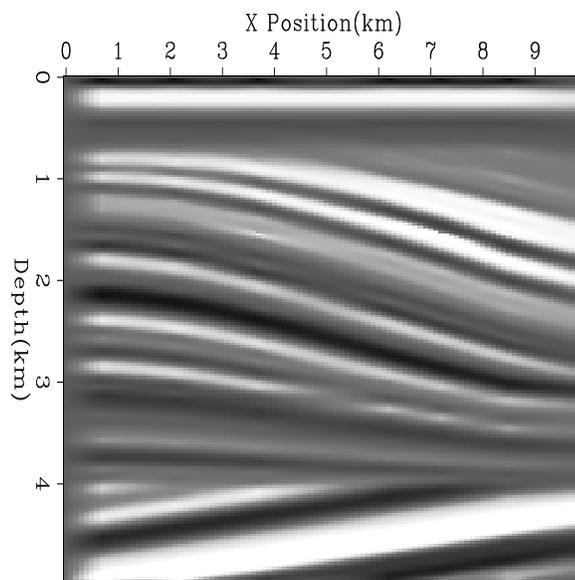


Figure 6: Left panel are four reflectors chosen to represent our *a priori* information. The right panel is interpolated slope calculated from the reflectors that will form the basis of our steering filter. `bob3-qdome-refs` [ER]

Figure 7: The result of using our steering filter operator as regularizer to the missing data problem.

`bob3-qdome-reg-cont` [ER]



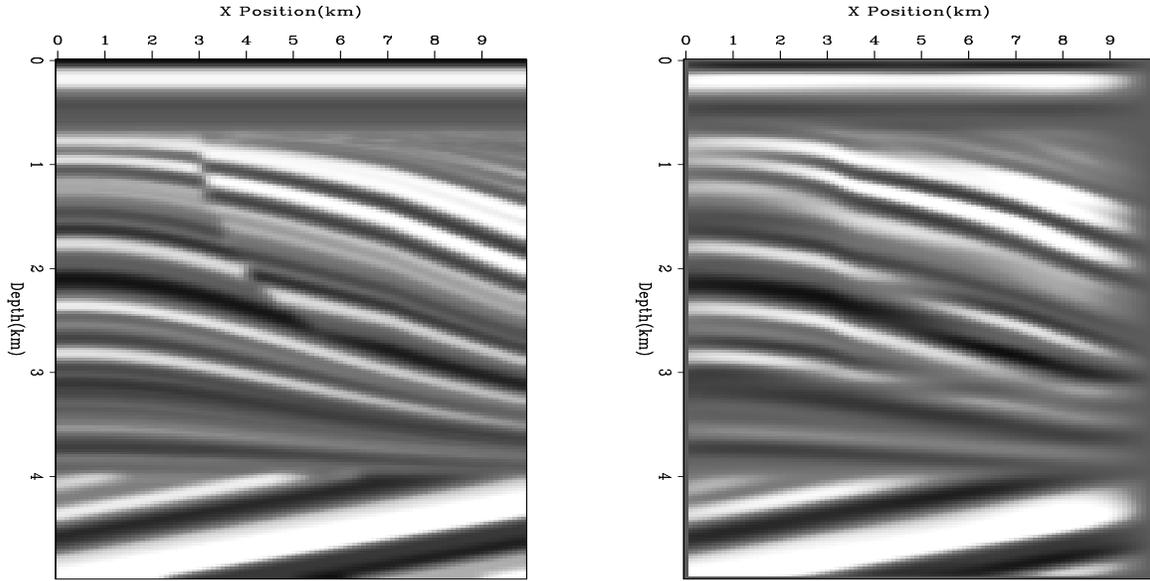


Figure 8: Left panel is a synthetic velocity model with the beds being offset by a listric fault. The right panel is the result of interpolating the model using steering filters. `bob3-fault-model` [ER]

ESTIMATING A NON-STATIONARY FILTER

Another option is to characterize our model in terms of PEFs rather than steering filters. Normally we estimate a PEF by solving

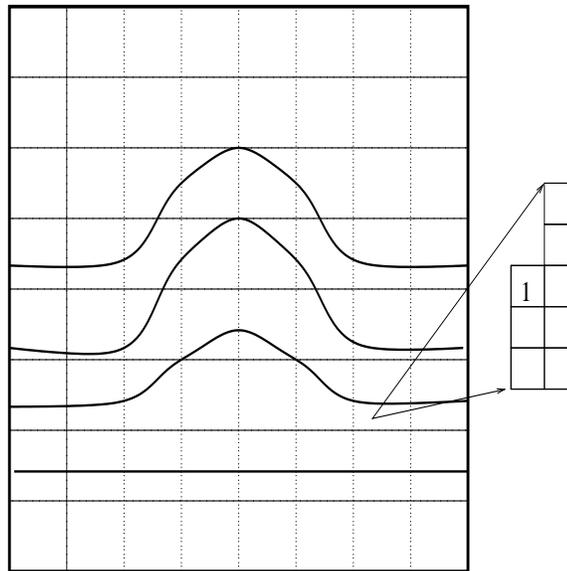
$$\mathbf{M}\mathbf{a} \approx \mathbf{0}, \quad (3)$$

where \mathbf{M} is convolution with a field that has the same properties as the model and \mathbf{a} is our PEF. The output of this convolution is white (Claerbout, 1992a). Therefore \mathbf{a} must have the inverse spectrum of the model. When the model varies continuously we must add a slight twist to our PEF estimation. Instead of breaking up our model space into regions where our stationary assumption is valid we are going to modify the PEF. Our PEF (\mathbf{a}) is now going to be composed of several different PEFs operating in micro-patches (Figure 9). With so many filters, and therefore filter coefficients, our filter estimation problem goes from being over-determined to under-determined. We can force the system to again be overdetermined by adding a regularization equation to our original filter estimation fitting goals,

$$\begin{aligned} \mathbf{0} &\approx \mathbf{M}\mathbf{a} \\ \mathbf{0} &\approx \epsilon\mathbf{F}\mathbf{a} \end{aligned} \quad (4)$$

where the regularization operator \mathbf{F} smooths the filter coefficients (Clapp et al., 1999).

Figure 9: Non-stationary PEF construction. The model is broken up into micro-patches. Each micro-patch has its own PEF. `bob3-patch` [NR]



MISSING DATA

To see the power of a non-stationary filter to characterize model covariance let's return to the fault model missing data problem (Figure 8). For our interpolation we will use the known model as the basis for our PEF, and have a micro-patch size of one sample in x and z . The resulting interpolation, Figure 10 isn't quite as high-frequency as our initial model and we have a minimal amount of continuation of the layers over the fault, but we generally do an excellent job recovering the model.

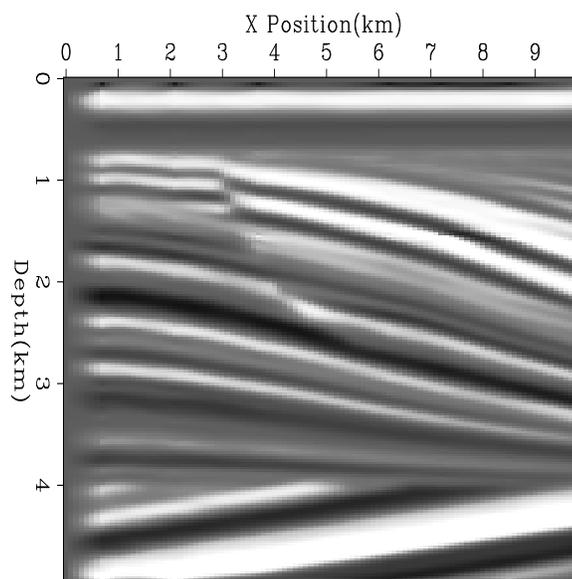


Figure 10: The result of finding non-stationary PEF using the correct model and then applying it to the missing data problem. The returned image is almost exact and does a better job with bed discontinuities than Figure 8. `bob3-fault.cont` [ER]

TOMOGRAPHY

Review

The next step is to see how well our non-stationary filter regularizes a tomography problem. I constructed a synthetic anticline model (Figure 11) with six reflectors, one above the anticline, four within the anticline, and one flat reflector representing basement rock. For added difficulty, there is a low velocity layer between the second and third reflector. The model was used to do acoustic wave modeling, with the resulting dataset having 32 meter CMP spacing and 80 offsets spaced 64 meters apart. If we use as our initial estimation of the slowness, an $s(z)$ function from outside the anticline, the reflectors are pulled up due to using too low a velocity within the anticline (Figure 12). Following the methodology of Clapp and Biondi (1999) we

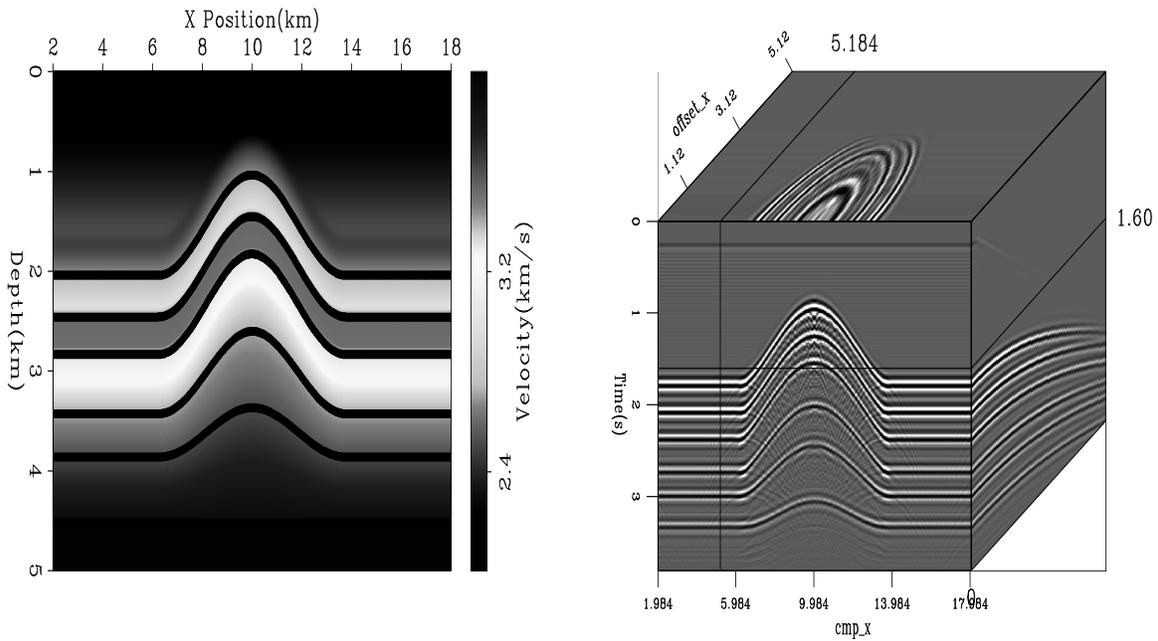


Figure 11: Left panel is the synthetic velocity model with six reflectors spanning the anticline. The right panel shows the data generated from this model. `bob3-synth-model` [ER]

will begin by considering a regularized tomography problem. We will linearize around an initial slowness estimate and find a linear operator in the vertical traveltimes domain \mathbf{T} between our change in slowness Δs and our change in traveltimes $\Delta \mathbf{t}$. We will write a set of fitting goals,

$$\begin{aligned} \Delta \mathbf{t} &\approx \mathbf{T} \Delta s \\ \mathbf{0} &\approx \epsilon \mathbf{A} \Delta s, \end{aligned} \quad (5)$$

where (\mathbf{A}) is our steering filter operator. However, these fitting goals don't accurately describe what we really want. Our steering filters are based on our desired slowness rather than change

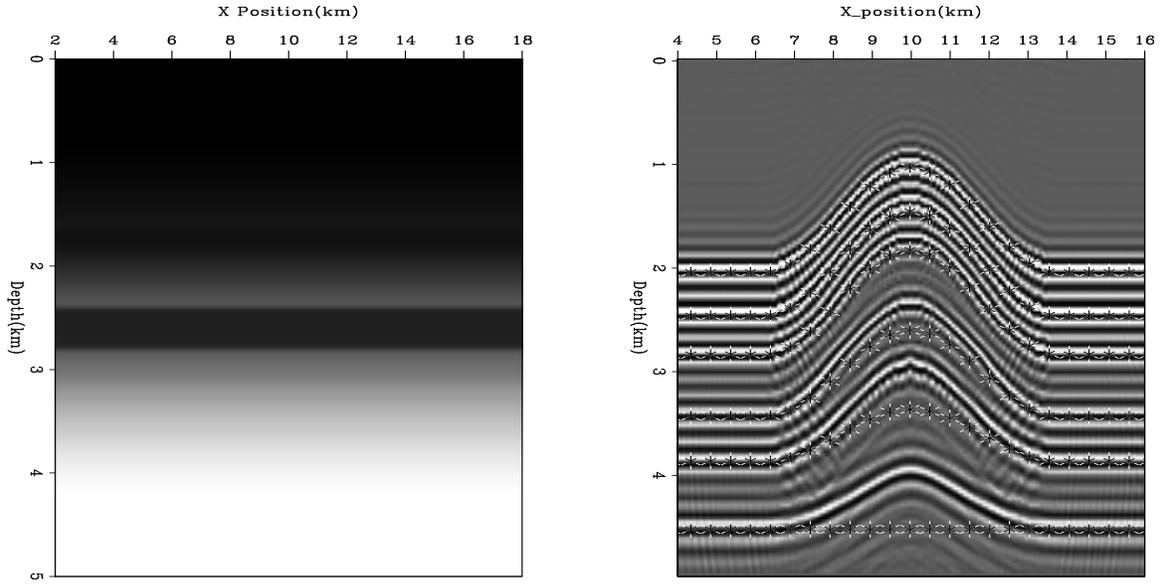


Figure 12: Left panel is our initial guess at the velocity function, the right panel shows the zero offset ray parameter reflector position using this migration velocity. The correct reflector positions are shown as ‘*’. Note that reflectors are significantly mispositioned. `bob3-mig0` [ER]

of slowness. With this fact in mind, we can rewrite our second fitting goal as:

$$\mathbf{0} \approx \mathbf{A}(\mathbf{s}_0 + \Delta\mathbf{s}) \quad (6)$$

$$-\epsilon\mathbf{A}\mathbf{s}_0 \approx \epsilon\mathbf{A}\Delta\mathbf{s}. \quad (7)$$

Our second fitting goal cannot be strictly defined as regularization but we can do a preconditioning substitution:

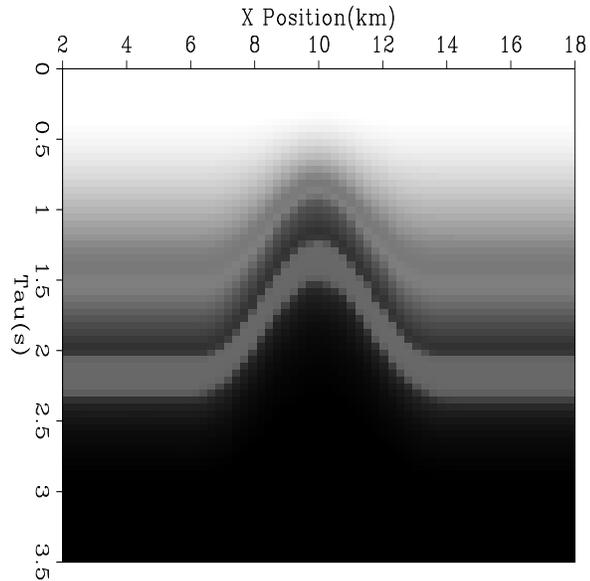
$$\begin{aligned} \Delta\mathbf{t} &\approx \mathbf{T}\mathbf{A}^{-1}\mathbf{p} \\ -\epsilon\mathbf{A}\mathbf{s}_0 &\approx \epsilon\mathbf{I}\mathbf{p}. \end{aligned} \quad (8)$$

Warping

There is one aspect of using non-stationary PEFs that I have glossed over to this point: the requirement that we have field with similar statistics to estimate the PEF from. Viewed one way this is a significant weakness to the approach, viewed another it can be seen as a useful feature. One of the largest problems in seismic imaging is how to put the geologist’s conception of geology into the geophysicist’s inversion problem. The non-stationary PEF can be estimated from the geologist’s model. Therefore, our regularization operator directly incorporates the geologist’s conception of the velocity structure into the velocity estimation.

For this simple model, we will use as our conception of geology a vertically warped version of our initial velocity estimate. We measure how much a migrated reflector positions varies from our initial flat estimate. Figure 13 shows our initial estimate.

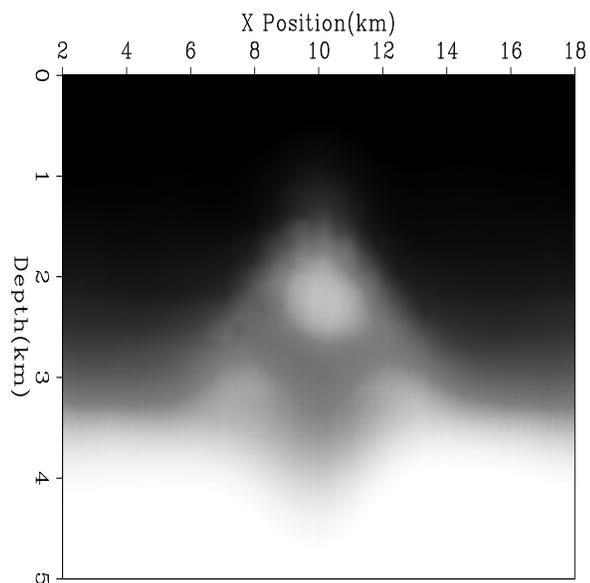
Figure 13: Warped velocity model used to estimate non-stationary filters from. `bob3-warp` [ER]



Results

If we do three non-linear iterations of tomography using fitting goals (8) each time, we get Figure 14 as our velocity estimate. The velocity estimate does a good job recovering the anticline shape. However, it doesn't do a good job recovering the low velocity layer. Migrating with Figure 14 we get Figure 15. Overlaid on top of the migrated image are the correct reflector positions. Overall we did a good job positioning the reflectors. In addition, we can see that we have little residual moveout in our CRP gathers (Figure 15) but they are relatively flat.

Figure 14: Velocity model after three non-linear iterations of non-stationary filter regularization. `bob3-vel3` [CR]



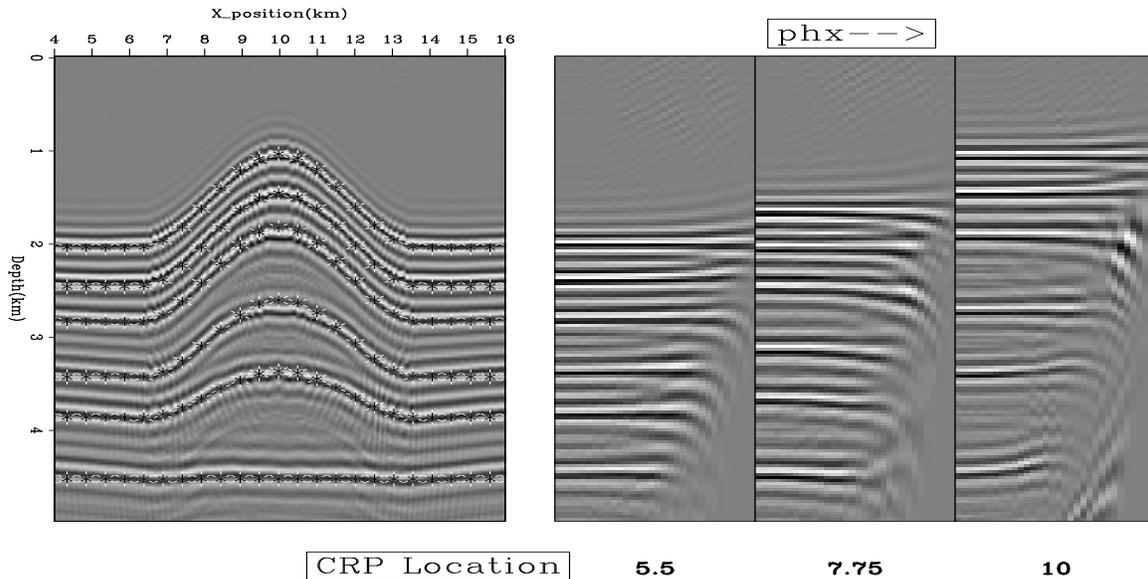


Figure 15: Migration result using Figure 14 as the velocity model. The left panel is the zero angle image overlaid by the correct reflector positions. The right three panels shows CRP gathers at six, eight, and ten kilometers. `bob3-res.vel3.from0` [CR]

CONCLUSIONS

Non-stationary PEFs are an effective interpolator, especially when the model has discontinuities, for missing data problems and an effective regularizer for tomography problems. The weakness of the approach, the requirement that we have field to estimate the PEF from, can be turned into a strength by allowing a geologist's conception of the velocity to be directly encoded into the regularization. A more illuminating test of the non-stationary regularization of tomography would be on a model with sharp, rather than smooth, boundaries.

REFERENCES

- Brown, M., 1998, Horizon refinement by synthesis of seismic and well log data: SEP-97, 193–206.
- Claerbout, J. F., 1990, Nonlinear problems: SEP-65, 229–240.
- Claerbout, J. F., 1992a, Earth Soundings Analysis: Processing versus Inversion: Blackwell Scientific Publications.
- Claerbout, J. F., 1992b, Nonstationarity and conjugacy: Utilities for data patch work: SEP-73, 391–400.
- Claerbout, J., 1999, Geophysical estimation by example: Environmental soundings image enhancement: Stanford Exploration Project, <http://sepwww.stanford.edu/sep/prof/>.

- Clapp, R. G., and Biondi, B. L., 1998, Regularizing time tomography with steering filters: SEP-97, 137-146.
- Clapp, R. G., and Biondi, B., 1999, Preconditioning tau tomography with geologic constraints: SEP-100, 35-50.
- Clapp, R. G., and Biondi, B. L., 2000, Tau tomography with steering filters: 2-D field data example: SEP-103, 1-19.
- Clapp, R. G., and Brown, M., 1999, Applying sep's latest tricks to the multiple suppression problem: SEP-102, 91-100.
- Clapp, R. G., and Brown, M., 2000, $(t - x)$ domain, pattern-based multiple separation: SEP-103, 201-210.
- Clapp, R. G., Fomel, S., and Claerbout, J., 1997, Solution steering with space-variant filters: SEP-95, 27-42.
- Clapp, R. G., Sava, P., and Claerbout, J. F., 1998, Interval velocity estimation with a null-space: SEP-97, 147-156.
- Clapp, R. G., Fomel, S., Crawley, S., and Claerbout, J. F., 1999, Directional smoothing of non-stationary filters: SEP-100, 197-209.
- Crawley, S., Clapp, R., and Claerbout, J., 1999, Interpolation with smoothly nonstationary prediction-error filters: 69th Ann. Internat. Meeting, Soc. Expl. Geophys., Expanded Abstracts, 1154-1157.
- Crawley, S., 1998, Shot interpolation for Radon multiple suppression: 68th Ann. Internat. Meeting, Soc. Expl. Geophys., Expanded Abstracts, 1238-1241.
- Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection data: Geophysics, **51**, no. 10, 1893-1903.
- Tarantola, A., 1987, Inverse problem theory: Elsevier.

Short Note

Traveltime sensitivity kernels: Banana-doughnuts or just plain bananas?

James Rickett¹

INTRODUCTION

Estimating an accurate velocity function is one of the most critical steps in building an accurate seismic depth image of the subsurface. In areas with significant structural complexity, one-dimensional updating schemes become unstable, and more robust algorithms are needed. Reflection tomography both in the premigrated (Bishop et al., 1985) and postmigrated domains (Stork, 1992; Kosloff et al., 1996) bring the powerful technologies of geophysical inversion theory to bear on the problem.

Unfortunately, however, inversion methods can be limited by the accuracy of their forward modeling operators, and most practical implementations of traveltime tomography are based on ray-theory, which assumes a high frequency wave, propagating through a smoothly varying velocity field, perhaps interrupted with a few discrete interfaces. Real world wave-propagation is much more complicated than this, and the failure of ray-based methods to adequately model wave propagation through complex media is fueling interest in “wave-equation” migration algorithms that both accurately model finite-frequency effects, and are practical for large 3-D datasets. As a direct consequence, finite-frequency velocity analysis and tomography algorithms are also becoming an important area of research (Woodward, 1992; Biondi and Sava, 1999).

Recent work in the global seismology community (Marquering et al., 1998, 1999) is drawing attention to a non-intuitive observation first made by Woodward (1992), that in the weak-scattering limit, finite-frequency traveltimes have zero-sensitivity to velocity perturbations along the geometric ray-path. This short-note aims to explore and explain this non-intuitive observation.

¹email: james@sep.stanford.edu

THEORY

A generic discrete linear inverse problem may be written as

$$\mathbf{d} = \mathbf{A} \mathbf{m} \quad (1)$$

where $\mathbf{d} = (d_1 \ d_2 \ \dots)^T$ is the known data vector, $\mathbf{m} = (m_1 \ m_2 \ \dots)^T$ is the unknown model vector, and \mathbf{A} represents the linear relationship between them. A natural question to ask is: which parts of the model influence a given observed data-point? The answer is that the row of matrix, \mathbf{A} , corresponding to the data-point of interest will be non-zero where that point in model space influences the data-value. Rows of \mathbf{A} may therefore be thought of as sensitivity kernels, describing which points in model space are sensed by a given data-point.

For a generic linearized travelttime tomography problem, travelttime perturbations, $\delta\mathbf{T}$, are related to slowness perturbations, $\delta\mathbf{S}$, through a linear system,

$$\delta\mathbf{T} = \mathbf{A} \delta\mathbf{S}. \quad (2)$$

The form of the sensitivity kernels depend on the the modeling operator, \mathbf{A} .

Under the ray-approximation, travelttime for a given ray, T , is calculated by integrating slowness along the ray-path,

$$T = \int_{\text{ray}} s(\mathbf{x}) \, dl. \quad (3)$$

Assuming that the ray-path is insensitive to a small slowness perturbation, the perturbation in travelttime is given by the path integral of the slowness perturbation along the ray,

$$\delta T = \int_{\text{ray}} \delta s(\mathbf{x}) \, dl. \quad (4)$$

Since travelttime perturbations given by equation (4) are insensitive to slowness perturbations anywhere off the geometric ray-path, the sensitivity kernel is identically zero everywhere in space, except along the ray-path where it is constant. The implication for ray-based travelttime tomography is that travelttime perturbations should be back-projected purely along the ray-path.

We are interested in more accurate tomographic systems of the form of equation (3), that model the effects of finite-frequency wave-propagation more accurately than simple ray-theory. Once we have such an operator, the first question to ask is: what do the rows look like?

Born travelttime sensitivity

One approach to building a linear finite-frequency travelttime operator is to apply the first-order Born approximation, to obtain a linear relationship between slowness perturbation, δS ,

and wavefield perturbation, δU ,

$$\delta \mathbf{U} = \mathbf{B} \delta \mathbf{S}. \quad (5)$$

The Born operator, \mathbf{B} , is a discrete implementation of equation (A-7), which is described in the Appendix.

Traveltime perturbations may then be calculated from the wavefield perturbation through a (linear) picking operator, \mathbf{C} , such that

$$\delta \mathbf{T} = \mathbf{C} \delta \mathbf{U} = \mathbf{CB} \delta \mathbf{S} \quad (6)$$

where \mathbf{C} is a (linearized) picking operator, and a function of the background wavefield, U_0 .

Cross-correlating the total wavefield, $U(t)$, with $U_0(t)$, provides a way of measuring their relative time-shift, δT . Marquering et al. (1999) uses this to provide the following explicit linear relationship between δT and $\delta U(t)$,

$$\delta T = \frac{\int_{t_1}^{t_2} \dot{U}(t) \delta U(t) dt}{\int_{t_1}^{t_2} \ddot{U}(t) U(t) dt}, \quad (7)$$

where dots denote differentiation with respect to t , and t_1 and t_2 define a temporal window around the event of interest. Equation (7) is only valid for small time-shifts, $\delta T \ll \lambda_{s_0}$.

Rytov traveltime sensitivity

The first Rytov approximation (or the phase-field linearization method, as it is also known) provides a linear relationship between the slowness and complex phase perturbations.

$$\delta \Psi = \mathbf{R} \delta \mathbf{S}, \quad (8)$$

where $\Psi = \exp(\mathbf{U})$, and the Rytov operator, \mathbf{R} , is a discrete implementation of equation (A-10), which is also described in Appendix A.

Traveltime is related to the complex phase by the equation, $\Im(\delta \psi) = \omega \delta t$. For a band-limited arrival with amplitude spectrum, $F(\omega)$, traveltime perturbation can be calculated simply by summing over frequency (Woodward, 1992),

$$\delta \mathbf{T} = \sum_{\omega} \frac{F(\omega)}{\omega} \Im(\delta \Psi) = \sum_{\omega} \frac{F(\omega)}{\omega} \Im(\mathbf{R} \delta \mathbf{S}). \quad (9)$$

Of the two approximations, several authors (Beydoun and Tarantola, 1988; Woodward, 1989) note that the Born approximation is the better choice for modeling reflected waves, while the Rytov approximation is better for transmitted waves. Differences tend to zero, however, as the scattering becomes small.

KERNELS COMPARED

This section contains images of traveltimes kernels computed numerically for a simple model that may be encountered in a reflection tomography problem. The source is situated at the surface, and the receiver (known reflection point) is located at a depth of 1.8 km in the subsurface. The background velocity model, $v_0(z) = 1/s_0(z)$, is a linear function of depth with $v_0(0) = 1.5 \text{ km s}^{-1}$, and $\frac{dv_0}{dz} = 0.8 \text{ s}^{-1}$. I chose a linear velocity function since Green's functions can be computed on-the-fly with rapid two-point ray-tracing.

Figure 1 shows the ray-theoretical traveltimes sensitivity kernel: zero except along the geometric ray-path.

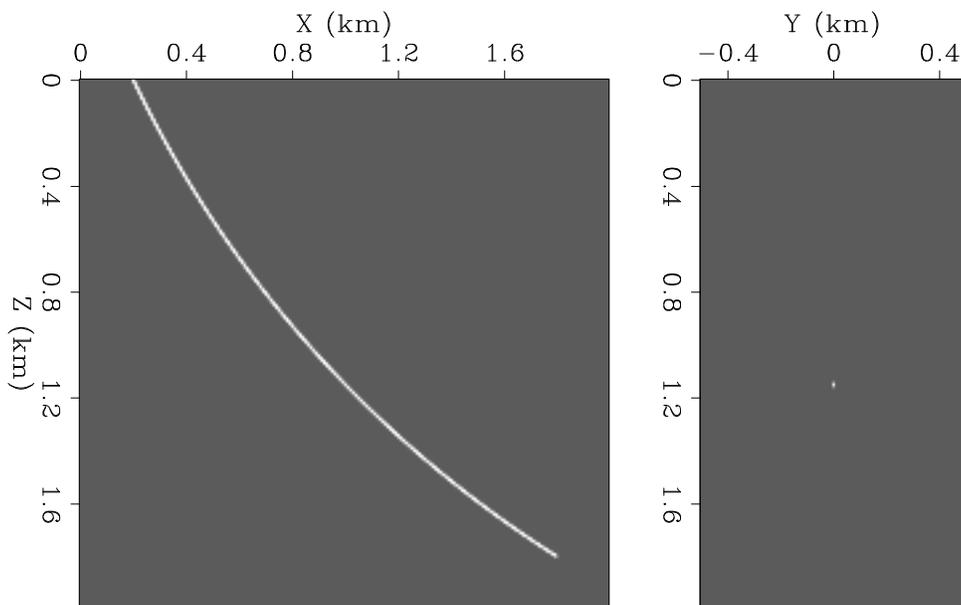


Figure 1: Traveltimes sensitivity kernel for ray-based tomography in a linear $v(z)$ model. The kernel is zero everywhere *except* along geometric ray-path. Right panel shows a cross-section at $X = 1 \text{ km}$. `james3-RayKernel` [ER]

Figures 2 and 3 show first Rytov traveltimes sensitivity kernels for 30 Hz and 120 Hz wavelets respectively. The important features of these kernels are identical to the features of kernels that Marquering et al. (1999) obtained for teleseismic $S - H$ wave scattering, and to Woodward's (1992) band-limited wave-paths. They have the appearance of a hollow banana: that is appearing as a banana if visualized in the plane of propagation, but as a doughnut on a cross-section perpendicular to the ray. Somewhat counter-intuitively, this suggests that traveltimes have zero sensitivity to small velocity perturbations along the geometric raypath. Fortunately, however, as the frequency of the seismic wavelet increases, the bananas become thinner, and approach the ray-theoretical kernels in the high-frequency limit. Parenthetically, it is also worth noticing that the width of the bananas increases with depth as the velocity (and seismic wavelength) increases.

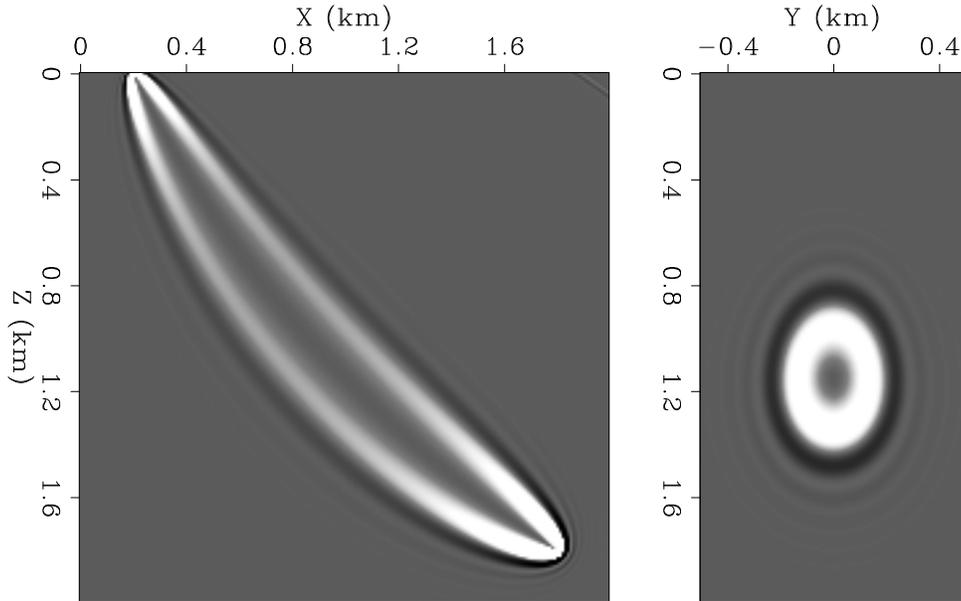


Figure 2: Rytov traveltime sensitivity kernel for 30 Hz wavelet in a linear $v(z)$ model. The kernel is zero along geometric ray-path. Right panel shows a cross-section at $X = 1$ km. [james3-BananaPancake8](#) [ER]

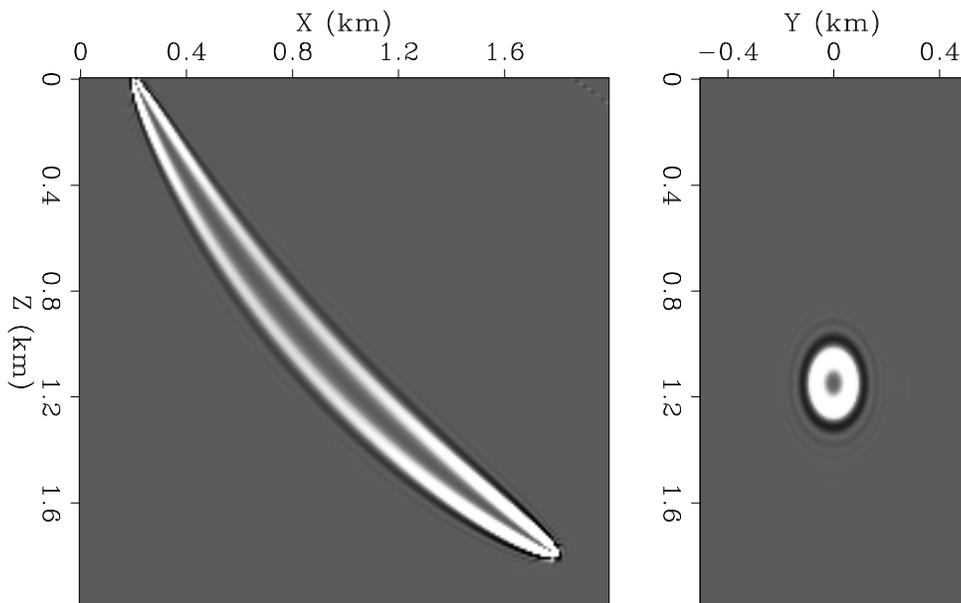


Figure 3: Rytov traveltime sensitivity kernel for 120 Hz wavelet in a linear $v(z)$ model. The kernel is zero along geometric ray-path. Right panel shows a cross-section at $X = 1$ km. [james3-BananaPancake2](#) [ER]

I do not show the first-Born kernels here, since, in appearance, they are identical to the Rytov kernels shown in Figures 2 and 3.

THE BANANA-DOUGHNUT PARADOX

The important paradox is not the apparent contradiction between ray-theoretical and finite-frequency sensitivity kernels, since they are compatible in the high-frequency limit. Instead, the paradox is how do you reconcile the zero-sensitivity along the ray-path with your intuitive understanding of wave propagation?

A first potential resolution to the paradox is that the wavefront healing removes any effects of a slowness perturbation. This alone is a somewhat unsatisfactory explanation since it does not explain why traveltimes are sensitive to slowness perturbations just off the geometric ray-path.

A second potential resolution is that the hollowness of the banana is simply an artifact of modeling procedure. This is partially true. Both Born and Rytov are single scattering approximations, and a single scatterer located on the geometric ray-path may only contribute energy in-phase with the direct arrival. In contrast, if there are two scatterers on the geometric ray-path traveltimes may be affected. However, just because the paradox may appear to be an artifact of the modeling procedure does not mean it is not a real phenomenon. In the weak scattering limit, traveltimes will indeed be insensitive to a slowness perturbation situated on the geometric ray-path.

CONCLUSIONS: DOES IT MATTER?

Practitioners of traveltime tomography typically understand the shortcomings of ray-theory; although they realize using “fat-rays” would be better, they smooth the slowness model both explicitly and by regularizing during the inversion procedure. In practice, any shortcomings of traveltime tomography are unlikely to be caused by whether or not the fat-rays are hollow.

However, the null space of seismic tomography problems is typically huge. Smoothing and regularization are often done with very ad hoc procedures. Understanding the effects of finite-frequency through sensitivity kernels may lead to incorporating more physics during the regularization and improve tomography results.

REFERENCES

- Arfken, G., 1985, *Mathematical methods for physicists*: Academic Press Inc., San Diego, 3rd edition.
- Beydoun, W. B., and Tarantola, A., 1988, First Born and Rytov approximations: Modeling and inversion conditions in a canonical example: *J. Acoust. Soc. Am.*, **83**, no. 3.

- Biondi, B., and Sava, P., 1999, Wave-equation migration velocity analysis: 69th Ann. Internat. Meeting, Soc. Expl. Geophys., 1723–1726.
- Bishop, T. N., Bube, K. P., Cutler, R. T., Langan, R. T., Love, P. L., Resnick, J. R., Shuey, R. T., Spindler, D. A., and Wyld, H. W., 1985, Tomographic determination of velocity and depth in laterally varying media: *Geophysics*, **50**, no. 06, 903–923.
- Kosloff, D., Sherwood, J., Koren, Z., MacHet, E., and Falkovitz, Y., 1996, Velocity and interface depth determination by tomography of depth migrated gathers: *Geophysics*, **61**, no. 5, 1511–1523.
- Marquering, H., Dahlen, F. A., and Nolet, G., 1998, Three-dimensional waveform sensitivity kernels: *Geophys. J. Int.*, **132**, 521–534.
- Marquering, H., Dahlen, F. A., and Nolet, G., 1999, Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox: *Geophys. J. Int.*, **137**, 805–815.
- Morse, P. M., and Feshbach, H., 1953, *Methods of theoretical physics*: McGraw-Hill, New York.
- Stork, C., 1992, Reflection tomography in the postmigrated domain: *Geophysics*, **57**, no. 5, 680–692.
- Woodward, M., 1989, A qualitative comparison of the first order Born and Rytov approximations: SEP-60, 203–214.
- Woodward, M. J., 1992, Wave-equation tomography: *Geophysics*, **57**, no. 1, 15–26.

APPENDIX A

BORN/RYTOV REVIEW

Modeling with the first-order Born (and Rytov) approximations [e.g. Beydoun and Tarantola (1988)] can be justified by the assumption that slowness heterogeneity in the earth exists on two separate scales: a smoothly-varying background, s_0 , within which the ray-approximation is valid, and weak higher-frequency perturbations, δs , that act to scatter the wavefield. The total slowness is given by the sum,

$$s(\mathbf{x}) = s_0(\mathbf{x}) + \delta s(\mathbf{x}). \quad (\text{A-1})$$

Similarly, the total wavefield, U , can be considered as the sum of a background wavefield, U_0 , and a scattered field, δU , so that

$$U(\mathbf{x}, \omega) = U_0(\mathbf{x}, \omega) + \delta U(\mathbf{x}, \omega), \quad (\text{A-2})$$

where U_0 satisfies the Helmholtz equation in the background medium,

$$[\nabla^2 + \omega^2 s_0^2(\mathbf{x})] U_0(\mathbf{x}, \omega) = 0, \quad (\text{A-3})$$

and the scattered wavefield is given by the (exact) non-linear integral equation (Morse and Feshbach, 1953),

$$\delta U(\mathbf{x}, \omega) = \frac{\omega^2}{4\pi} \int_V G_0(\mathbf{x}, \omega; \mathbf{x}') U(\mathbf{x}, \omega; \mathbf{x}') \delta s(\mathbf{x}') dV(\mathbf{x}'). \quad (\text{A-4})$$

In equation (A-4), G_0 is the Green's function for the Helmholtz equation in the background medium: i.e. it is a solution of the equation

$$[\nabla^2 + \omega^2 s_0^2(\mathbf{x})] G_0(\mathbf{x}, \omega; \mathbf{x}_s) = -4\pi \delta(\mathbf{x} - \mathbf{x}_s). \quad (\text{A-5})$$

Since the background medium is smooth, in this paper I use Green's functions of the form,

$$G_0(\mathbf{x}, \omega; \mathbf{x}_s) = A_0(\mathbf{x}, \mathbf{x}_s) e^{i\omega T_0(\mathbf{x}, \mathbf{x}_s)}. \quad (\text{A-6})$$

where A_0 and T_0 are ray-traced traveltimes and amplitudes respectively.

A Taylor series expansion of U about U_0 for small δs , results in the infinite Born series, which is a Neumann series solution (Arfken, 1985) to equation (A-4). The first term in the expansion is given below: it corresponds to the component of wavefield that interacts with scatters only once.

$$\delta U_{\text{Born}}(\mathbf{x}, \omega) = \frac{\omega^2}{4\pi} \int_V G_0(\mathbf{x}, \omega; \mathbf{x}') U_0(\mathbf{x}, \omega; \mathbf{x}') \delta s(\mathbf{x}') dV(\mathbf{x}'). \quad (\text{A-7})$$

The approximation implied by equation (A-7) is known as the first-order Born approximation. It provides a linear relationship between δU and δs , and it can be computed more easily than the full solution to equation (A-4).

The Rytov formalism starts by assuming the heterogeneity perturbs the phase of the scattered wavefield. The total field, $U = \exp(\psi)$, is therefore given by

$$U(\mathbf{x}, \omega) = U_0(\mathbf{x}, \omega) \exp(\delta\psi) = \exp(\psi_0 + \delta\psi). \quad (\text{A-8})$$

The linearization based on small $\delta\psi/\psi$ leads to the infinite Rytov series, on which the first term is given by

$$\delta\psi_{\text{Rytov}}(\mathbf{x}, \omega) = \frac{\delta U_{\text{Born}}(\mathbf{x}, \omega)}{U_0(\mathbf{x}, \omega)} \quad (\text{A-9})$$

$$= \frac{\omega^2}{4\pi U_0(\mathbf{x}, \omega)} \int_V G_0(\mathbf{x}, \omega; \mathbf{x}') U_0(\mathbf{x}, \omega; \mathbf{x}') \delta s(\mathbf{x}') dV(\mathbf{x}'). \quad (\text{A-10})$$

The approximation implied by equation (A-10) is known as the first-order Rytov approximation. It provides a linear relationship between $\delta\psi$ and δs .

Short Note

Research interest: Improving the velocity model

Daniel Rosales¹

INTRODUCTION

Travel time is perhaps one of the most important seismic parameter. It can be used to estimate properties of the subsurface. Velocity is one of these properties, and its importance has been increasing because it is the unique element that can convert time into depth. Seismic exploration has been evolving from time studies to depth studies because they map actual geology. Velocities can be also used to estimate petrophysical properties.

My goal is to find interval velocities from seismic data. Seismic velocities were considered important only as a parameter for stacking seismic data (one of the most important computer process in the prospecting industry (Claerbout, 1985)) or for converting structure maps from time to depth.

There are many methods of building an accurate velocity model; one of these uses maximum stacking power. The velocity model built in this way yields it to one kind of seismic velocity: the root mean square velocity (V_{rms}). Interval velocity is sometimes derived from rms velocity. Processes such as prestack depth migration are very sensitive to errors in the velocity model. Because of this, prestack depth migration is also used as a tool to estimate an interval velocity models (Sava, 2000). Methodologies such as Residual Curvature Analysis (Al-Yahya, 1989) and Depth Focusing Analysis (Faye and Jeannot, 1986) are different ways of building a velocity model. Tomography is another technique of estimating velocities from seismic data (Clapp and Biondi, 2000).

We can note that there are many ways of estimating velocities from seismic data, and it is not easy to find a solution to improve such estimation.

In the following pages, I will introduce a method that incorporates well information and seismic data; this method can help us to improve the velocity model. It is not a new approach (Brown, 1998; Clapp et al., 1997), but I review those works and provide a basis on which we can build our ideas about the velocity model building process.

¹email: daniel@sep.stanford.edu

HOW TO IMPROVE A VELOCITY MODEL

Using inverse theory, the velocity estimation is usually an overdetermined problem. The use of all available sources of velocity information could help us to overcome this overdetermination. Seismic data provide velocity information (residual moveout analysis, dix equation). In addition Vertical seismic profiles give vertical traveltimes and well data provide major geological information (dip, velocities).

Seismic exploration is at a point where the fields are in very complex areas such as salt domes and overthrusts. Classically, the union of seismic information with surface geology has been used to improve the velocity model in an interpretative way, with the work of geophysicists and geologists combined to build a velocity model with an important geological component. This kind of work has been developed in such a way that now, the cooperation of both kinds of geoscientists is of primary importance in the continued development of the velocity model.

I think that a combination of seismic interval velocities along with well information (velocities, dips measurements) could bring in a more refined velocity model.

Problems involved

The introduction of well data in the velocity model demands some extra care because seismic data provide a relatively high horizontal resolution while well data provide a very narrow vertical resolution but a poor horizontal resolution. Well velocities are direct rock properties, estimated directly from the earth, while seismic velocities are indirect rock properties estimated from traveltimes. If the combination of seismic velocities and well velocities were possible, migration results would be favorably improved.

The problem is how to manage both velocities to get better seismic images. Well velocities may be interpolated to obtain a horizontal velocity model. This velocity model can be correlated with the velocity model from seismic data, resulting in a new velocity model from both kinds of information. The interpolation problem, may be solved using geostatistics. As a conclusive example, Lee and Xu (2000) show how geostatistics helped to improve the imaging of a geopressure zone in the Gulf of Mexico. Other interpolation techniques (Fomel, 1997) can be implemented in order to obtain a satisfactory *well velocity model*. Steering filters prove effective in interpolating velocities obtaining a horizontal velocity model from well data (Clapp et al., 1997, 1998; Clapp and Biondi, 1998), but it will be necessary to regularize well velocities in order to allow a satisfying the correlation between the velocity model from seismic and from well to produce good results. In the next section, I am going to discuss a method to obtain interval velocities from RMS velocities.

INTERVAL VELOCITY ESTIMATION

A basic daily problems in seismic processing, such as the estimation of interval velocities from RMS velocities, will be solved in this part.

The method used here was first introduced by Clapp et al. (1998). The method builds a velocity model from surface seismology while retaining the null-space. They start from fundamental concepts in Geophysical Estimation by Example (Claerbout, 1997) and define the simplest interval velocity estimation including the notion of null-space. Generally, Clapp et al. (1998) minimize interval velocities “wiggleness” where there are not good quality reflections.

In order to understand the method used in this part it is necessary to make some definitions (for further explanation the reader could refer to Clapp et al. (1998):

C as the matrix of causal integration, a lower triangular matrix of ones.

D as the matrix of causal differentiation, namely, $\mathbf{D} = \mathbf{C}^{-1}$.

u as a vector whose components range over the vertical traveltime depth τ , and whose component values contain the interval velocity squared v_{interval}^2 .

d as a data vector whose components range over the vertical travel time depth τ , and whose component values contain the scaled RMS velocity squared $\tau v_{\text{RMS}}^2 / \Delta\tau$ where $\tau / \Delta\tau$ is the index on the time axis.

The theoretical (squared) RMS velocity is defined by

$$\mathbf{Cu} = \mathbf{d}. \quad (1)$$

With imperfect data, our data fitting goal is to minimize the residual

$$\mathbf{0} \approx \mathbf{W}[\mathbf{Cu} - \mathbf{d}]. \quad (2)$$

To find the interval velocity where there is no data, we have the “model damping” goal to minimize the “wiggleness” **p** of the squared interval velocity **u**

$$\mathbf{0} \approx \mathbf{Du} = \mathbf{p}. \quad (3)$$

These two goals are preconditioned by changing the optimization variable from interval velocity squared **u** to its wiggleness **p**. Substituting $\mathbf{u} = \mathbf{Cp}$ gives the two fitting goals expressed as a function of wiggleness **p**

$$\mathbf{0} \approx \mathbf{W}'[\mathbf{C}^2\mathbf{p} - \mathbf{d}] \quad (4)$$

$$\mathbf{0} \approx \epsilon\mathbf{p}. \quad (5)$$

This method was tested on two synthetic CMP gathers and one real CMP gather from the Gulf of Mexico.

Synthetic 1

A synthetic interval velocity model was built and a CMP gather was modeled using a finite difference code (Figure 1)

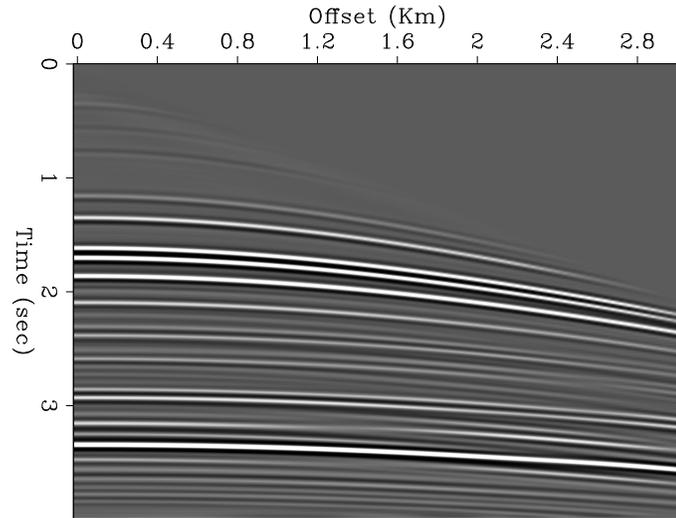


Figure 1: Synthetic Model.
[daniel1-synt-bob] [CR]

For this gather, a velocity analysis is performed to obtain the RMS velocity \mathbf{d} (equation 1). Figure 2 shows the rms velocity curve picked from the data, and Figure 3 shows a comparison of the interval velocity \mathbf{u} obtained by the methodology described above, the RMS velocity \mathbf{d} obtained from the velocity analysis, and the remodeled RMS velocity ($\mathbf{d}_{\text{remod}} = \mathbf{C}\mathbf{u}$).

This comparison shows that the remodeled RMS velocity is similar to the picked RMS velocity. This similarity prove that the method used works well.

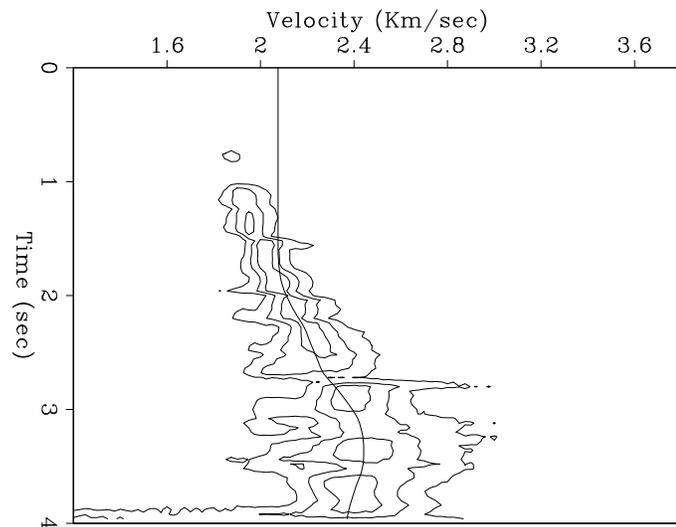


Figure 2: Velocity analysis for the first synthetic example. The curve corresponds to the picked RMS velocity [daniel1-scan-bob] [CR]

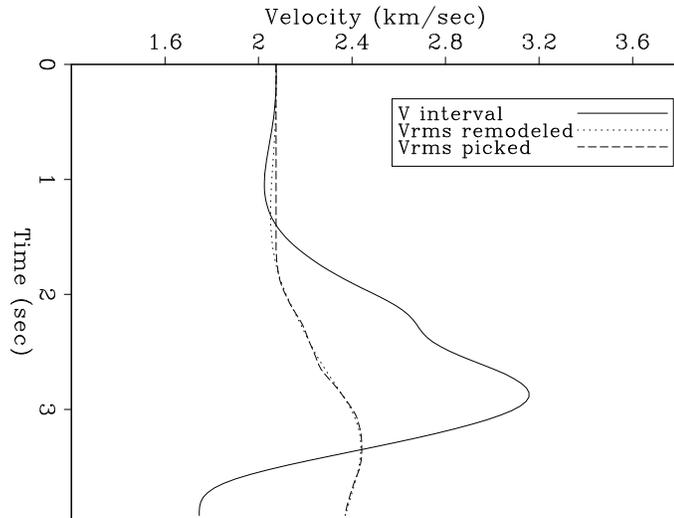


Figure 3: Comparison of interval and RMS velocities. daniel1-compar-bob [CR]

Synthetic 2

Some petrophysical parameters were carefully chosen to generate a synthetic model that simulates a geopressure zone (Dutta, 1987; Mavko et al., 1998). Figure 4 displays the velocity model used. At 3 km, it is possible to see a strong velocity change that identifies the geopressure zone. This velocity anomaly is visible in Figure 5 at 2.3 sec where a polarity inversion occurs.

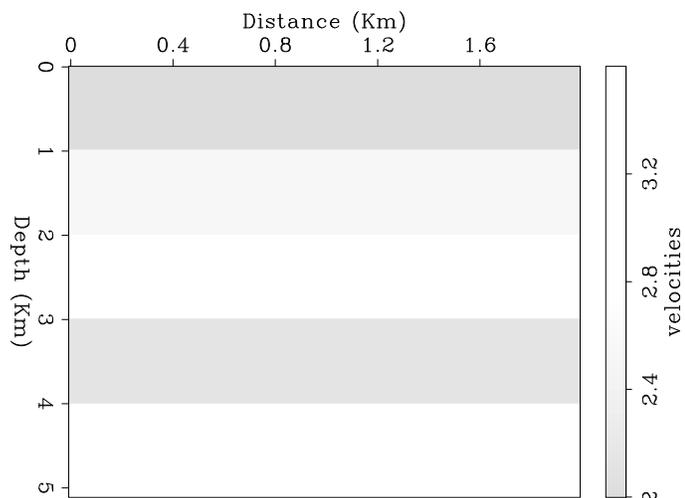


Figure 4: Depth velocity model for a geopressure zone. daniel1-model [CR]

Figure 6 shows the velocity analysis of this synthetic and the picked RMS. Finally, Figure 7 exhibit a comparison of the interval velocity obtained with the inversion method, the picked RMS velocity and the remodeled RMS velocity.

Figure 5: Synthetic gather simulating a geopressure zone.
 daniel1-synt-inv [CR]

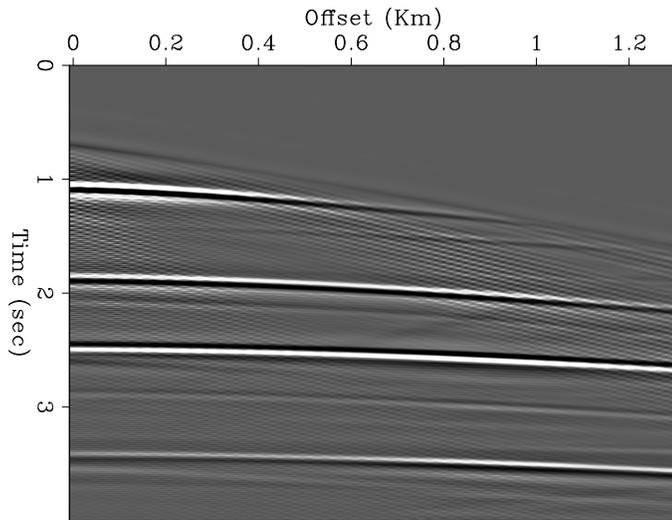


Figure 6: Velocity analysis for the second synthetic example. The curve is the picked RMS velocity.
 daniel1-scan-inv [CR]

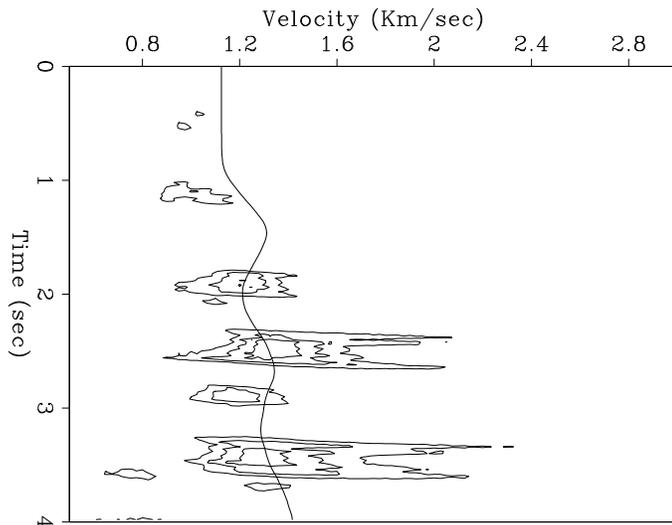
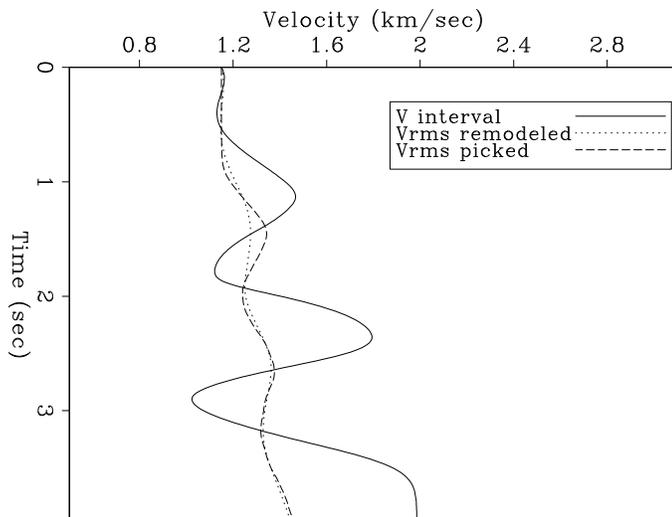


Figure 7: Comparison of interval velocity and RMS velocities.
 daniel1-compar-inv [CR]



Real Data

A CMP gather from the Gulf of Mexico (Figure 8) is used. The velocity analysis is displayed in Figure 9, with its corresponding V_{rms} picked in a velocity fairway. The comparison of the interval velocity obtained with the inversion method, the picked RMS velocity and the remodeled V_{rms} is illustrated in Figure 10.

Figure 8: CMP gather from the Gulf of Mexico. `daniel1-data` [CR]

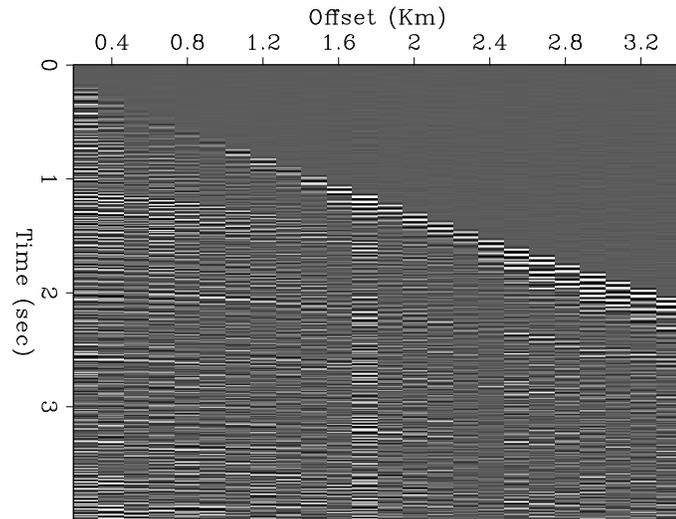
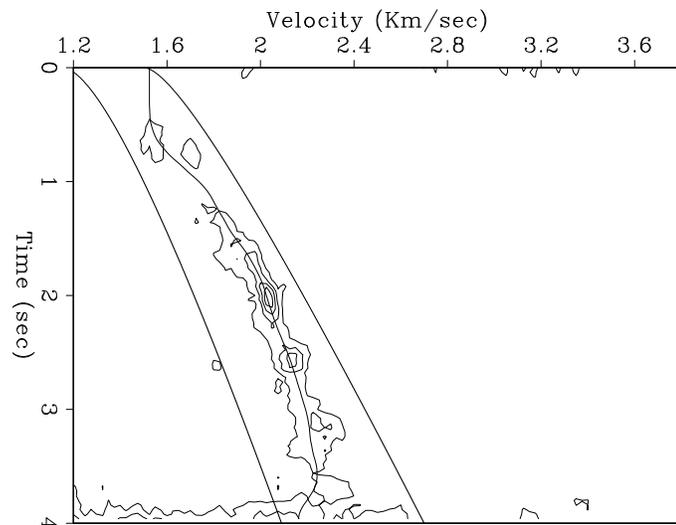


Figure 9: Velocity analysis for the real data. The picking for V_{rms} is done in a velocity fairway. `daniel1-scan-data` [CR]



FUTURE WORK

One of the next step will be to incorporate and interpolate dip data obtained from dip meters. This data converted into time could be utilized as a new parameter in the velocity estimation problem. Finally, interpolation of well velocities will give us a velocity model. That can be correlated with the seismic velocity model. This way, we hope to have a more accurate velocity model for future analysis.

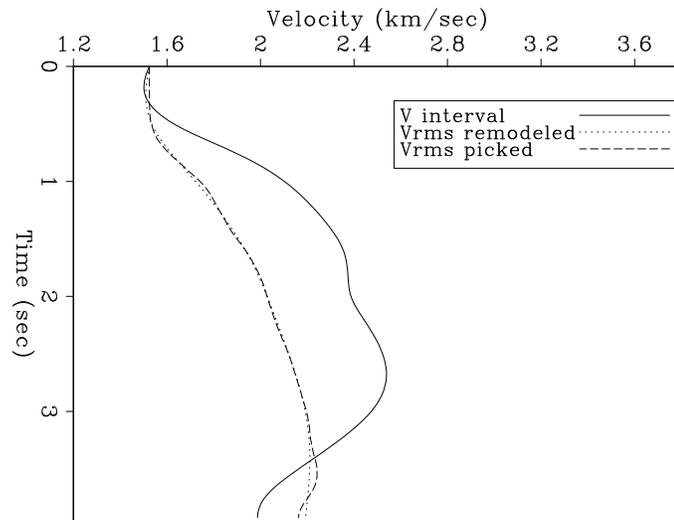


Figure 10: Comparison of interval velocity and RMS velocities.
daniel1-compar-data [CR]

ACKNOWLEDGMENTS

I thank Biondo Biondi, James Rickett, for their useful comments and Carmen Mora for her computational support. I deeply thank Antoine Guitton for his thorough proofreading and very useful comments.

REFERENCES

- Al-Yahya, K. M., 1989, Velocity analysis by iterative profile migration: *Geophysics*, **54**, no. 6, 718–729.
- Brown, M., 1998, Horizon refinement by synthesis of seismic and well log data: *SEP-97*, 193–206.
- Claerbout, J. F., 1985, *Imaging the Earth's Interior*: Blackwell Scientific Publications.
- Claerbout, J. F., 1997, *Geophysical exploration mapping: Environmental soundings image enhancement*: Stanford Exploration Project.
- Clapp, R. G., and Biondi, B. L., 1998, Regularizing time tomography with steering filters: *SEP-97*, 137–146.
- Clapp, R. G., and Biondi, B. L., 2000, Tau tomography with steering filters: 2-D field data example: *SEP-103*, 1–19.
- Clapp, R. G., Fomel, S., and Claerbout, J., 1997, Solution steering with space-variant filters: *SEP-95*, 27–42.
- Clapp, R. G., Sava, P., and Claerbout, J. F., 1998, Interval velocity estimation with a null-space: *SEP-97*, 147–156.

Dutta, N. C., Ed., 1987, Geopressure, Geophysics Reprint Series No. 7 Society of Exploration Geophysicists, Tulsa.

Faye, J. P., and Jeannot, J. P., 1986, Prestack migration velocities from focusing depth analysis: 56th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, Session:S7.6.

Fomel, S., 1997, On the general theory of data interpolation: SEP-**94**, 165–179.

Lee, W., and Xu, W., 2000, 3D geostatistical velocity modeling: Salt imaging in a geopressed environment: The Leading Edge, **19**, no. 01, 32–36.

Mavko, G., Mukerji, T., and Dvorkin, J., 1998, The rock physics handbook: Cambridge.

Sava, P., 2000, Variable-velocity prestack Stolt residual migration with application to a North Sea dataset: SEP-**103**, 147–157.

Stable wide-angle Fourier-finite difference downward extrapolation of 3-D wavefields

*Biondo Biondi*¹

ABSTRACT

I derive an unconditionally stable implicit finite-difference operator that corrects the constant-velocity phase shift operator for lateral velocity variations. My method is based on the Fourier-Finite Difference (FFD) method first proposed by Ristow and Ruhl (1994). Contrary to previous results, my correction operator is stable even when the reference velocity is higher than the medium velocity. Because of this additional capability, after the correction step I can apply a frequency-dependent interpolation that significantly reduces: the residual phase error after correction, the frequency dispersion caused by the discretization of the Laplacian operator, and the azimuthal anisotropy caused by splitting. Tests on zero-offset data from the SEG-EAGE salt data set show that the proposed method improves the imaging of a fault reflection with respect to a similar interpolation scheme that employs extended split-step to adapt to lateral velocity variations.

INTRODUCTION

As 3-D prestack wave-equation imaging is becoming practically possible (Biondi and Palacharla, 1996; Mosher et al., 1997), we need robust, efficient, and accurate methods to downward continue 3-D wavefields. In particular, wide-angle methods are crucial for prestack imaging because at least one of the paths connecting the image point in the subsurface to the source/receiver locations at the surface is likely to propagate at a wide angle.

Fourier methods, such as phase shift (Gazdag, 1978), handle wide-angle propagation efficiently and accurately, but only for vertically layered media. In contrast, finite-difference methods can easily handle lateral velocity variations, but are not efficient for wide-angle propagation. A natural strategy is thus to combine a Fourier method with a finite-difference method to derive an extrapolation method that enjoys the strengths of both. This is not a new idea, and indeed both methods that were first proposed to adapt Fourier methods to lateral velocity variations, Phase Shift Plus Interpolation (PSPI) (Gazdag and Sguazzero, 1984) and split-step (Stoffa et al., 1990), can be interpreted as being zero-order finite-difference corrections to a phase shift extrapolator. Ristow and Ruhl (1994) first proposed a genuinely finite-difference correction to phase shift, which they dubbed Fourier-Finite Difference (FFD), that employs im-

¹email: biondo@sep.stanford.edu

implicit finite differences (Claerbout, 1985) to handle lateral velocity variations. Pseudo-screen propagators (Jin et al., 1998), generalized-screen propagators (Le Rousseau and de Hoop, 1998), and local Born-Fourier migration (Huang et al., 1999) are related methods that have been recently proposed in the literature.

In the first part of this paper, I show that the FFD correction is more accurate than the implicit finite-difference implementation of the first order Born-Fourier and pseudo-screen corrections. Because the computational complexity of the three methods is comparable, the FFD correction is more attractive than the others, and it is the focus of my paper. Unfortunately, my first attempts at applying the original FFD method failed because of numerical instability arisen when the velocity model had sharp discontinuities, such as the one encountered at the boundaries of a salt body. Stability is a necessary condition for a migration method to be practically useful.

The stable FFD correction that I present in this paper overcomes the instability problems related to the original FFD method. To derive a stable version of the FFD correction I adapted the bullet-proofing theory developed by Godfrey et al. (1979) and Brown (1979) for the 45-degree equation. The bullet-proofed FFD correction is unconditionally stable for arbitrary variations in the medium velocity *and* in the reference velocity. Further, I demonstrate that it is unconditionally stable when the medium velocity is either higher *or* lower than the reference velocity. This is a useful result, and differs with a statement in Ristow and Ruhl's paper, that asserts their method to be unstable when the medium velocity is lower than the reference velocity.

The stability of the new FFD correction, even when the reference velocity is higher than the medium velocity and it is laterally varying, makes it a suitable building block for the construction of a new wide-angle downward continuation algorithm that is efficient and accurate in 3-D. At each depth step, the wavefield is propagated with n reference velocities using phase shift, where n is determined according to the maximum propagation angle that we need to propagate accurately. Then the n reference wavefields are combined to create two wavefields: one for which the reference velocity is lower than the medium velocity, the other one for which the reference velocity is higher than the medium velocity. A stable FFD correction is applied to both wavefields, and the corrected wavefields are linearly interpolated with frequency-dependent weights. The frequency-dependent interpolation enables a significant reduction of the frequency dispersion introduced by the discretization of the Laplacian operator in the implicit finite difference step.

In 3-D, the FFD corrections can be efficiently applied by splitting, or possibly by helix-transform methods (Rickett, 2000). However, the proposed algorithm suffers much less from azimuthal anisotropy caused by splitting than the original FFD method. The phase variations as a function of azimuth have opposite behavior when the differences between the reference velocity and medium velocity have opposite signs. Therefore, these phase variations tend to cancel each other when the two wavefields are interpolated after the correction.

Because Fourier-finite differences methods and interpolation are both fundamental components of the new method, I will refer to it as Fourier-Finite Difference Plus Interpolation method, or FFDPI.

PSEUDO-SCREEN VS. FFD CORRECTION

The simplest way to derive the pseudo-screen, or local Born-Fourier, correction is to expand the square root of the one-way wave equation in a Taylor series and stop at the first order. The wavefield downward continued at depth $z + \Delta z$ is then computed from the wavefield at depth z using the following approximation,

$$P_{z+\Delta z} = P_z e^{ik_z^v \Delta z} \approx P_z e^{ik_z^{v_r} \Delta z + i \frac{dk_z}{ds} \Delta s \Delta z}, \quad (1)$$

where the vertical wavenumber k_z^v for the medium velocity v is given by the well known single square root equation

$$k_z^v = \frac{\omega}{v} \sqrt{1 - \frac{v^2 k_m^2}{\omega^2}}, \quad (2)$$

and the first derivative of the square root with respect to the slowness s evaluated at the reference velocity v_r is

$$\frac{dk_z}{ds} = \frac{\omega}{\sqrt{1 - \frac{v_r^2 k_m^2}{\omega^2}}}. \quad (3)$$

If we employ a finite-difference method to apply the correction term, we need to approximate the square root in equation (3). A reasonable approximation is achieved by using a Taylor expansion, as follows,

$$\frac{dk_z}{ds} \approx \omega \left[1 + \frac{v_r^2 X^2}{2} + \frac{3v_r^4 X^4}{8} + \dots \right], \quad (4)$$

where $X = k_m/\omega$. A computationally more efficient way is to use Muir's continued fraction approximation (1981), that is about as accurate as the second order Taylor expansion but includes only second order terms for the spatial derivatives (i.e. X^2 instead of X^2 and X^4); that is,

$$\frac{dk_z}{ds} \approx \omega \left[1 + \frac{\frac{v_r^2 X^2}{2}}{1 - \frac{3v_r^2 X^2}{4}} \right]. \quad (5)$$

Notice that the first term of the correction in both equations (4) and (5) corresponds to the split-step correction term (Stoffa et al., 1990).

Figure 1 demonstrates the accuracy improvement gained by including the second term in equation (5). It compares the phase curves obtained after applying the first term in equation (5) (split step) and after applying both terms (pseudo screen). The medium velocity v is equal to 2 km/s, and two reference velocities are assumed: one 10% lower than the medium velocity (1.8 km/s), the other one 10% higher than the medium velocity (2.2 km/s).

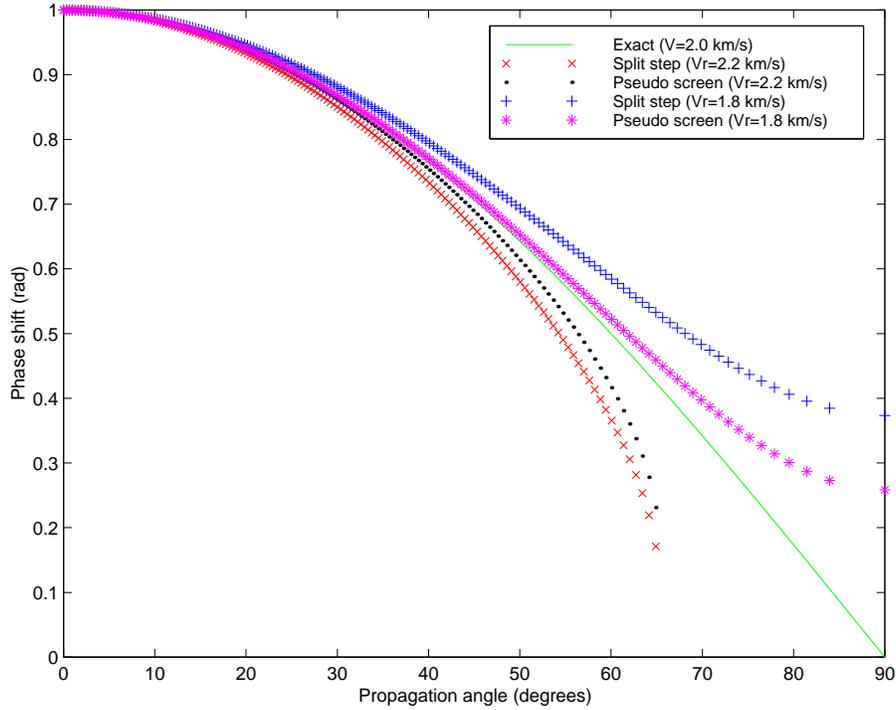


Figure 1: Phase curves for: (-) Exact with $v = 2$ km/s, (x) Split step with $v = 2.2$ km/s, (.) Pseudo screen with $v = 2.2$ km/s, (+) Split step with $v = 1.8$ km/s, (*) Pseudo screen with $v = 2.2$ km/s, `biondo1-born` [CR]

FFD correction

The FFD correction achieves better accuracy than the pseudo-screen correction because it is based on a direct expansion of the difference between the square root evaluated at the medium velocity v and the square root evaluated at the reference velocity v_r , instead of being based on the expansion of the square root around the reference velocity. The downward continued wavefield is approximated as

$$P_{z+\Delta z} = P_z e^{ik_z^v \Delta z} = P_z e^{ik_z^{v_r} \Delta z + i \frac{\Delta k_z}{\Delta s} \Delta s \Delta z}, \quad (6)$$

where the Taylor series of the correction term is now

$$\frac{\Delta k_z}{\Delta s} \approx \omega \left[1 + \frac{v_r v X^2}{2} + \frac{v_r v (v_r^2 + v^2 + v_r v) X^4}{8} + \dots \right], \quad (7)$$

and the continued fraction approximation of the correction term is

$$\frac{\Delta k_z}{\Delta s} \approx \omega \left[1 + \frac{\frac{v_r v X^2}{2}}{1 - \frac{(v_r^2 + v^2 + v_r v) X^2}{4}} \right]. \quad (8)$$

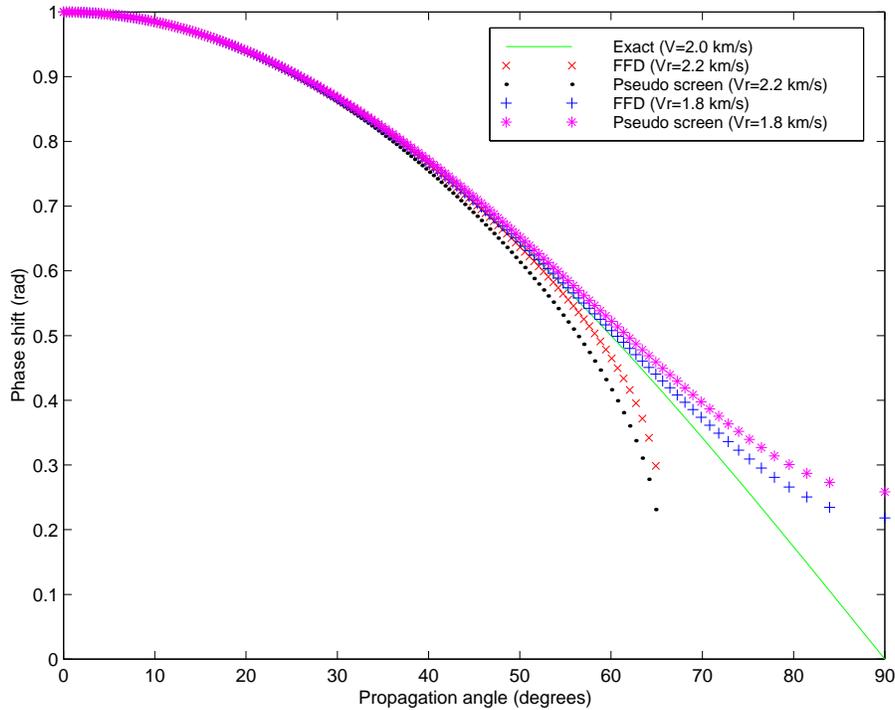


Figure 2: Phase curves for: (-) Exact with $v = 2$ km/s, (x) FFD with $v = 2.2$ km/s, (.) Pseudo screen with $v = 2.2$ km/s, (+) FFD with $v = 1.8$ km/s, (*) Pseudo screen with $v = 2.2$ km/s, biondo1-ristow [CR]

Notice that both equations (7) and (8) respectively reduce to equations (4) and (5) if $v = v_r$. Therefore, at the limit when the difference between the reference velocity and the medium velocity is small the two correction terms are equivalent, but they differ for larger corrections. The superiority of the FFD correction is demonstrated in Figure 2. It compares the phase curves obtained after applying the pseudo-screen correction [equation (5)] and the FFD correction [equation (8)]. As in Figure 1, the medium velocity v is equal to 2 km/s, and two reference velocities are assumed: one 10% lower than the medium velocity (1.8 km/s), the other one 10% higher than the medium velocity (2.2 km/s).

Figures 3–5 show the impulse responses associated with the phase curves shown in Figure 2. The maximum frequency in the data is 42 Hz and the spatial sampling is 10 m in both directions. Figure 3 shows the exact impulse response for the medium velocity equal to 2 km/s. Figure 4 shows the impulse response with reference velocity equal to 1.8 km/s and FFD correction. Figure 5 shows the impulse response with reference velocity equal to 2.2 km/s and FFD correction. Notice the frequency dispersion in the image obtained applying the FFD correction. These artifacts are caused by the discretization errors of the horizontal second derivatives in X^2 . The phase curves shown in Figure 2 neglect this approximation, and thus they represent the effective phase shift for zero-frequency data. Also notice that the frequency dispersion is in the opposite directions for opposite signs of the velocity correction.

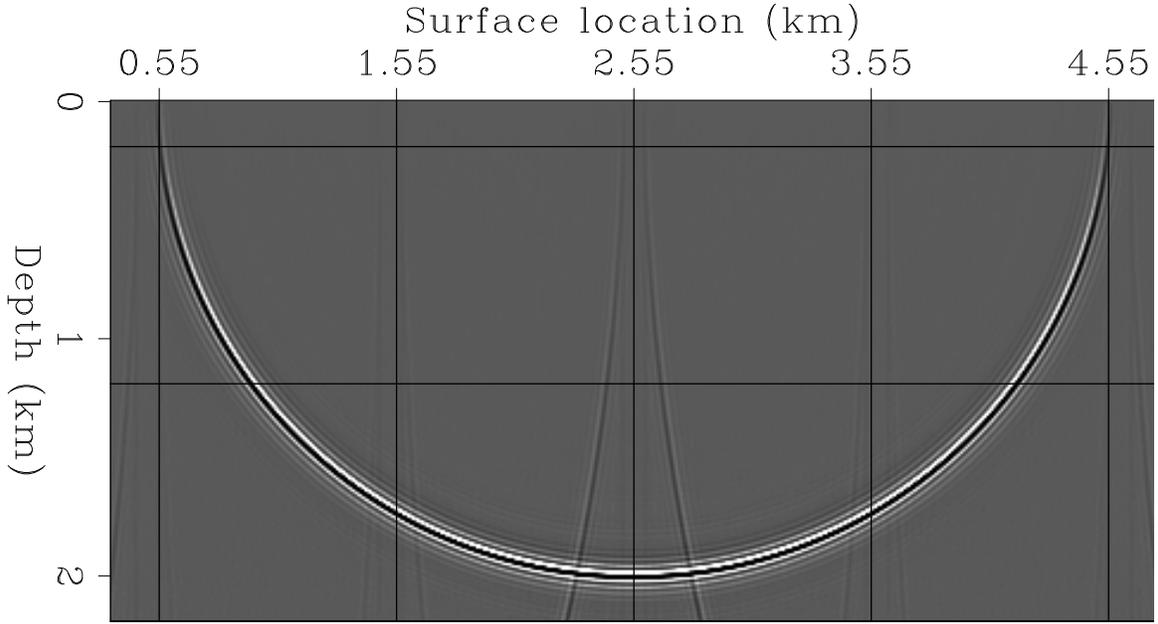


Figure 3: Impulse response for the medium velocity equal to 2 km/s. biondo1-Line-rist-zero [ER]

STABLE FFD CORRECTION

An implicit finite-difference implementation using a Crank-Nicolson scheme of the FFD correction as expressed in equation (8) is stable for smooth velocity variations. But numerical instability may develop when there are sharp discontinuities in the velocity field. An example of this situation is shown in Figures 6 and 7. The slowness function (Figure 6) has a sharp negative step, and a random behavior within the low slowness region. The impulse response computed applying the original FFD correction is shown in Figure 7, and clearly illustrates the problem. Notice that the image was clipped at the 70th percentile before plotting, to allow the “shadow” of the familiar circular impulse response to be visible in the plot.

In constant velocity the correction operator is unitary (all-pass filter) because its eigenvalues have zero imaginary part. Numerical instability originates when variations in the velocities terms multiplying the second derivative [$v_r v X^2$ and $(v_r^2 + v^2 + v_r v) X^2$] causes the imaginary part to become different from zero. To assure that this does not happen we first rewrite equation (8) as

$$\frac{\Delta k_z}{\Delta s} \Delta s \approx \omega \left\{ \frac{(v_r - v)}{v_r v} + \left[\frac{2(v_r - v)}{v_r^2 + v^2 + v_r v} \right] \frac{\frac{(v_r^2 + v^2 + v_r v) X^2}{4}}{1 - \frac{(v_r^2 + v^2 + v_r v) X^2}{4}} \right\}. \quad (9)$$

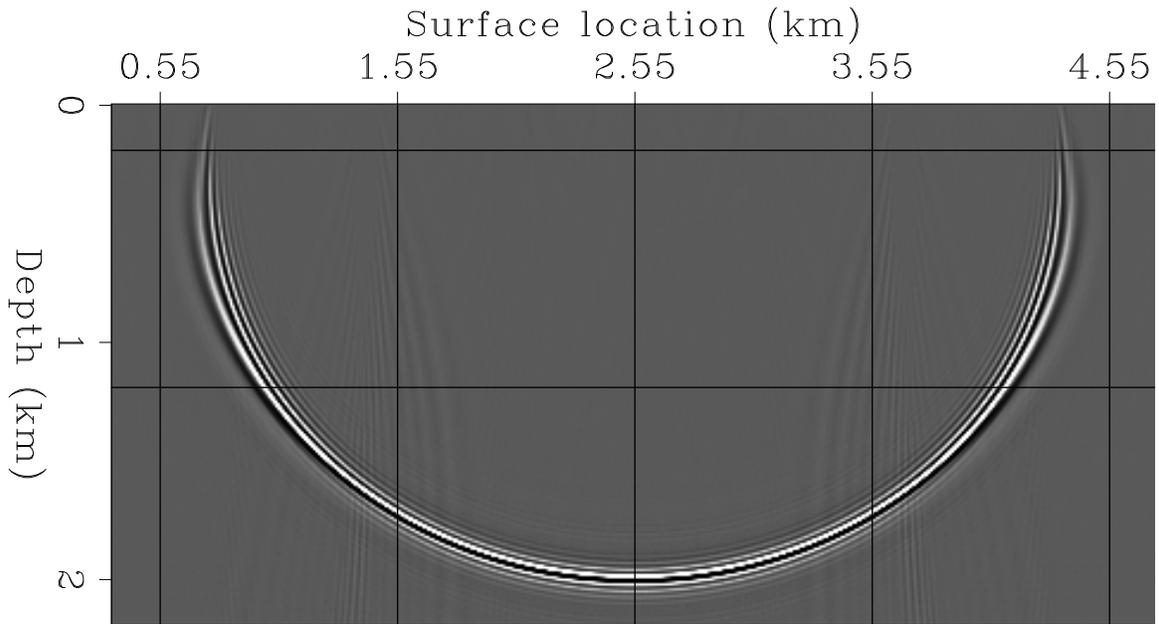


Figure 4: Impulse response with reference velocity equal to 1.8 km/s and FFD correction.
`biondo1-Line-rist-pos` [ER]

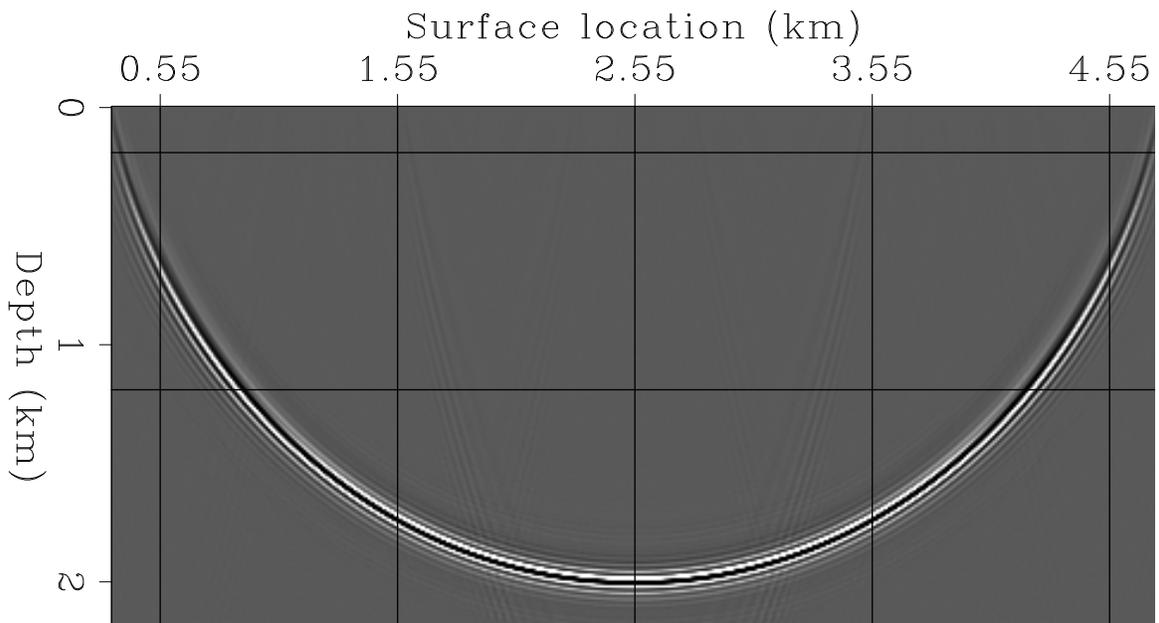


Figure 5: Impulse response with reference velocity equal to 2.2 km/s and FFD correction.
`biondo1-Line-rist-neg` [ER]

Then we rewrite $(v_r^2 + v^2 + v_r v) X^2$ as the product of a matrix with its adjoint, that is,

$$\Sigma' X^2 \Sigma = \frac{1}{\omega^2 \Delta x^2} \Sigma' D' D \Sigma = \frac{1}{\omega^2 \Delta x^2} \Sigma' T \Sigma = \frac{1}{\omega^2 \Delta x^2} \Sigma' \begin{bmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ 0 & -1 & 2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 2 \end{bmatrix} \Sigma \quad (10)$$

where

$$\Sigma = \frac{1}{2} \begin{bmatrix} \sqrt{1v^2 + 1v_r^2 + 1v_1 v_r} \\ \vdots \\ \sqrt{i-1v^2 + i-1v_r^2 + i-1v_{i-1} v_r} \\ \sqrt{iv^2 + iv_r^2 + iv_i v_r} \\ \sqrt{i+1v^2 + i+1v_r^2 + i+1v_{i+1} v_r} \\ \vdots \\ \sqrt{nv^2 + nv_r^2 + nv_n v_r} \end{bmatrix}, \quad (11)$$

where ${}_i v_r$ and ${}_i v$ are respectively the reference velocity and medium velocity at the i grid horizontal location.

The matrix $\Sigma' X^2 \Sigma$ is now guaranteed to have real eigenvalues. Because both I and $\Sigma' X^2 \Sigma$ are normal matrices that can be diagonalized by the same similarity transformation (Brown, 1979), also the matrix $(I + \Sigma' X^2 \Sigma)^{-1} \Sigma' X^2 \Sigma$ is now guaranteed to have real eigenvalues.

In matrix notation equation (9) can be rewritten as

$$\frac{\Delta k_z}{\Delta s} \Delta s \approx \omega \left[\frac{(v_r - v)}{v_r v} + \text{sign}(v_r - v) \Delta \frac{\Sigma' X^2 \Sigma}{I + \Sigma' X^2 \Sigma} \right] \quad (12)$$

where

$$\Delta = 2 \text{Diag} \begin{bmatrix} \frac{|1v_r - 1v|}{1v^2 + 1v_r^2 + 1v_1 v_r} \\ \vdots \\ \frac{|i-1v_r - i-1v|}{i-1v^2 + i-1v_r^2 + i-1v_{i-1} v_r} \\ \frac{|iv_r - iv|}{iv^2 + iv_r^2 + iv_i v_r} \\ \frac{|i+1v_r - i+1v|}{i+1v^2 + i+1v_r^2 + i+1v_{i+1} v_r} \\ \vdots \\ \frac{|nv_r - nv|}{nv^2 + nv_r^2 + nv_n v_r} \end{bmatrix}. \quad (13)$$

I pulled out the sign of the velocity perturbations outside the diagonal matrix Δ to perform the next, and last, step of the bullet-proofing process. To demonstrate that the multiplication by Δ does not introduce instability I first recognize that multiplying the wavefield by the exponential of the second term in equation (12) is equivalent to solve the differential equation

$$\frac{d}{dz}P = i\omega \operatorname{sign}(v_r - v) \Delta \frac{\Sigma' X^2 \Sigma}{I + \Sigma' X^2 \Sigma} P. \quad (14)$$

Notice that the equivalence is true only if $\operatorname{sign}(v_r - v)$ is constant; that is, if the reference velocity v_r is always lower or always higher than the medium velocity v . Second, I perform the change of variable $P = \Delta^{\frac{1}{2}} Q$, and equation (14) becomes

$$\frac{d}{dz}Q = \Delta^{\frac{1}{2}} \left(i\omega \operatorname{sign}(v_r - v) \frac{\Sigma' X^2 \Sigma}{I + \Sigma' X^2 \Sigma} \right) \Delta^{\frac{1}{2}} Q = \Delta^{\frac{1}{2}} (i\omega S) \Delta^{\frac{1}{2}} Q. \quad (15)$$

The norm $\|Q\|_{\Delta^{-1}}$ is constant with depth because the eigenvalues of S are real, and it obeys the following differential equation

$$\frac{d}{dz} \|Q\|_{\Delta^{-1}}^2 = Q^* (i\omega S - i\omega S^*) Q = 0. \quad (16)$$

Equation (16) guarantees the stability of the new FFD correction, independently from the value of $\operatorname{sign}(v_r - v)$, as long it is constant. In theory $(v_r - v)$ should be never equal to zero for the previous arguments to be strictly valid to avoid singularities in the norm $\|Q\|_{\Delta^{-1}}$. In practice, however, I never had to enforce this condition, though it would be easy to do so.

The reference velocity v_r and the medium velocity v can be interchanged at will in the previous development without changing the stability conditions. Therefore, the stable FFD correction is not only stable in presence of sharp discontinuities in the medium velocity, but also in presence of sharp discontinuities in the reference velocity. This property is exploited in the next section to design an efficient and accurate interpolation scheme.

Equation (14) can be solved using a Crank-Nicolson scheme and the wavefield at depth $z + \Delta z$ computed as

$$P_{z+\Delta z} = \frac{1 + i\operatorname{sign}(v_r - v)\omega \frac{\Delta z}{2} \Delta \frac{\Sigma' X^2 \Sigma}{I + \Sigma' X^2 \Sigma}}{1 - i\operatorname{sign}(v_r - v)\omega \frac{\Delta z}{2} \Delta \frac{\Sigma' X^2 \Sigma}{I + \Sigma' X^2 \Sigma}} P_z. \quad (17)$$

Figure 8, shows the same impulse response as in Figure 7, but computed using the stable FFD correction. In this case no numerical instability is encountered and the wavefield propagates without problems through the region with random slowness perturbations.

In 3-D the stable FFD correction could be applied using a 2-D version of the Crank-Nicolson method summarized in equation (17). However, the solution of the linear system implied by equation (17) would be expensive. A splitting algorithm (Jakubowicz and Levin, 1983) can be employed to reduce computational costs. The stability of a splitting algorithm derives directly from the analysis above, since it consists of the successive application of the FFD correction along the two horizontal coordinate axes. In the next section I discuss how the capability of using both positive and negative velocity corrections yields a significant improvement in accuracy of the splitting algorithm.

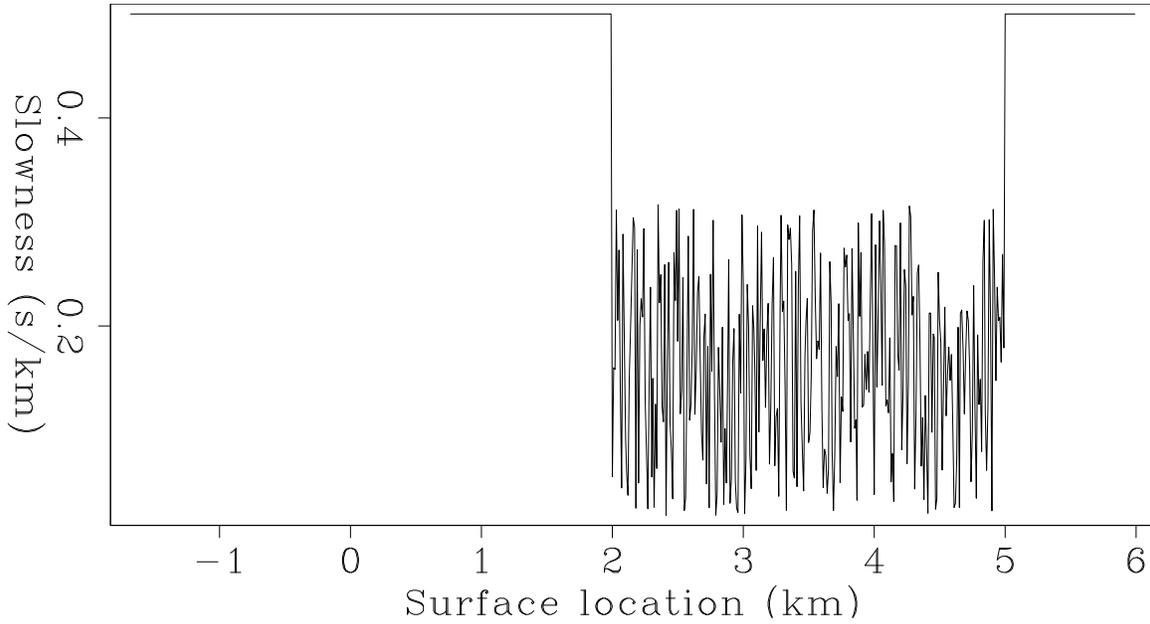


Figure 6: Depth slice through the slowness function that causes the original FFD correction to become unstable `biondo1-Slow-random-pos` [ER]

Boundary conditions

A necessary component of deriving a stable downward continuation scheme is to define stable boundary conditions. It is also desirable for the boundaries to be absorbing. Following Clayton and Engquist (1977), and Rothman and Thorson (1982), this goal can be easily accomplished by changing the values at the edges of the diagonal of T in equation (10) from 2 to $(1 + p_b)$, where $p_b = [p_r, i\text{sign}(v_r - v)|p_i|]$; that is substituting T with

$$T_b = \begin{bmatrix} (1 + p_b) & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ 0 & -1 & 2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & (1 + p_b) \end{bmatrix}. \quad (18)$$

The sign of the imaginary part of p_b determines whether the boundaries are absorbing or amplifying; therefore, the sign of $(v_r - v)$ must be constant to avoid instability to develop at the boundaries. The actual values of p_r and p_i determine the propagation angle of the incident wavefield that is most strongly attenuated. In practice, to assure both strong attenuation and weak reflection from the boundaries, I use boundary strips few sample wide, instead of a single boundary layer.

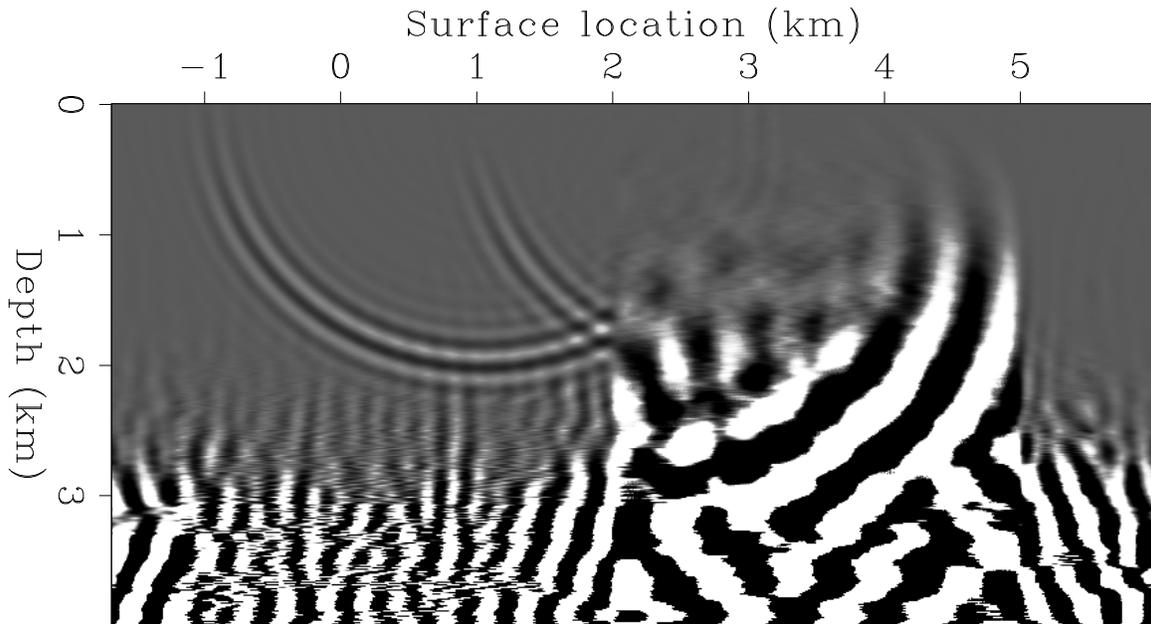


Figure 7: Impulse response computed with the original FFD correction and assuming the slowness function shown in Figure 6. The image was clipped at the 70th percentile before plotting, to allow the “shadow” of the familiar circular impulse response to be visible in the plot. `biondo1-Image-random-pos-rist2-1` [ER]

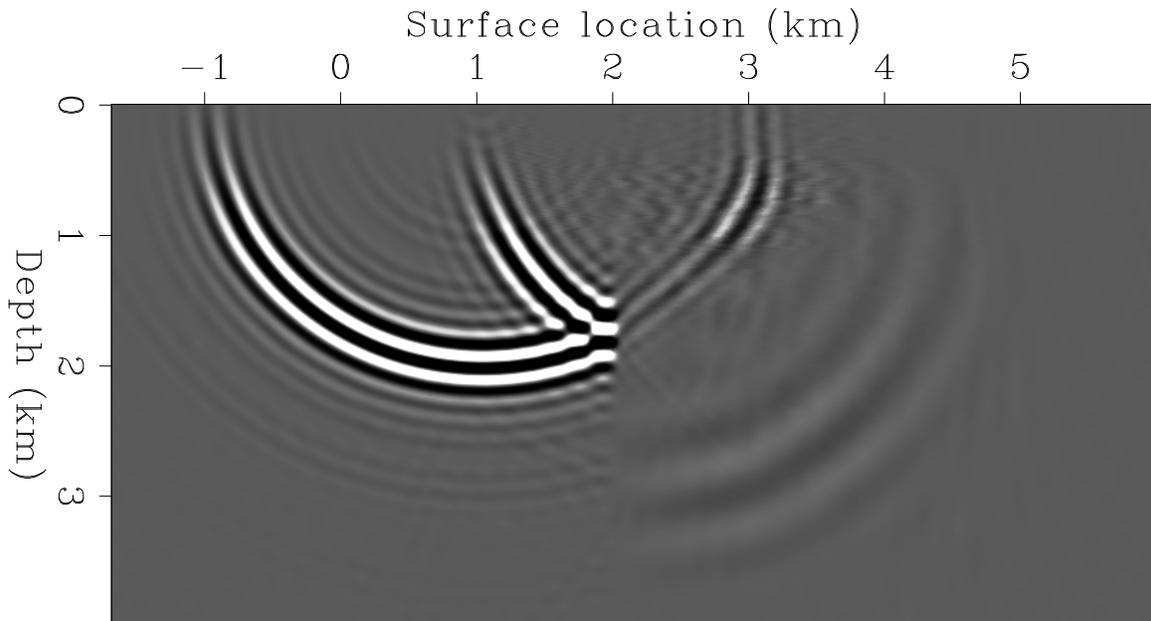


Figure 8: Impulse response computed with the stable FFD correction and assuming the slowness function shown in Figure 6 `biondo1-Image-random-pos-rist2-stable-1` [ER]

THE FOURIER-FINITE DIFFERENCE PLUS INTERPOLATION ALGORITHM

The stable FFD correction developed in the previous section has the right characteristics to be used as the main building block of an efficient and accurate wide-angle continuation algorithm. To achieve accuracy we can interpolate between wavefields that have been phase shifted using several reference velocities and corrected using the stable FFD method. In theory, arbitrary accuracy can be achieved by increasing the number of reference velocities. The structure of the algorithm is similar to the PSPI method, except that a wide-angle correction (FFD) is employed instead of a narrow-angle one (vertical shift). This change reduces the errors over the whole range of propagation angles.

Two results reached in the previous section are important for the definition of a stable and accurate interpolation scheme. First, the application of the FFD correction is stable independently from the sign of the velocity correction to be applied, as long as it is constant within the same correction step. This result enables a *linear interpolation* between a wavefield corresponding to reference velocities lower than the medium velocity and a wavefield corresponding to reference velocities higher than the medium velocity. Previously, because of the requirement for the reference velocity to be lower than the medium velocity, only a *nearest-neighborhood interpolation* was possible. Second, the reference velocity can vary at will laterally, without compromising the stability of the method. Because of these results, computations can be saved by applying the FFD correction only twice at each depth step. Once starting from a wavefield constructed from all the reference wavefields computed using a reference velocity lower than the medium velocity. The second time starting from a wavefield constructed from all the reference wavefields computed using a reference velocity higher than the medium velocity.

The algorithm outlined above can be described in more details as the sequence of the following steps.

1. Downward continue the data using phase shift for N_{v_r} reference velocities v_r^j , and compute the reference wavefields $P_{z+\Delta z}^{v_r^j}$ as

$$P_{z+\Delta z}^{v_r^j} = P_z e^{ik_z^{v_r^j} \Delta z} \quad \text{with } j = 1, \dots, j, \dots, N_{v_r}. \quad (19)$$

2. Define two reference velocity functions v_r^- and v_r^+ , that at every point are respectively equal to the reference velocity that is just lower and just higher than the medium velocity; that is,

$$v_r^- = \sum_{j=1}^{N_{v_r}} \delta^- v_r^j \quad \text{where } \delta^- = \begin{cases} 1 & \text{if } v_r^j \leq v < v_r^{j+1} \\ 0 & \text{elsewhere} \end{cases} \quad (20)$$

$$v_r^+ = \sum_{j=1}^{N_{v_r}} \delta^+ v_r^j \quad \text{where } \delta^+ = \begin{cases} 1 & \text{if } v_r^{j-1} \leq v < v_r^j \\ 0 & \text{elsewhere} \end{cases}. \quad (21)$$

$$(22)$$

3. Extract from the reference wavefields two wavefields corresponding to v_r^+ and v_r^- and correct them using the stable FFD method:

$$P_{z+\Delta z}^- = e^{i \frac{\Delta k_z}{\Delta s} \frac{(v_r^- - v)}{v_r^- v} \Delta z} \sum_{j=1}^{N_{v_r}} \delta^- P_{z+\Delta z}^{v_r^j} \quad (23)$$

$$P_{z+\Delta z}^+ = e^{i \frac{\Delta k_z}{\Delta s} \frac{(v_r^+ - v)}{v_r^+ v} \Delta z} \sum_{j=1}^{N_{v_r}} \delta^+ P_{z+\Delta z}^{v_r^j}. \quad (24)$$

4. Linearly interpolate the two corrected wavefield as

$$P_{z+\Delta z} = W^- P_{z+\Delta z}^- + W^+ P_{z+\Delta z}^+. \quad (25)$$

The interpolation weights can be chosen in order to zero the phase error for a given propagation angle as follows,

$$W^- = \begin{cases} \frac{k_z^v - k_z^{v_r^+} - \frac{\Delta k_z}{\Delta s} \frac{(v_r^+ - v)}{v_r^+ v}}{k_z^{v_r^-} + \frac{\Delta k_z}{\Delta s} \frac{(v_r^- - v)}{v_r^- v} - k_z^{v_r^+} - \frac{\Delta k_z}{\Delta s} \frac{(v_r^+ - v)}{v_r^+ v}} & \text{if } \Im(k_z^{v_r^+}) = 0 \\ 1 & \text{elsewhere} \end{cases} \quad (26)$$

$$W^+ = 1 - W^-.$$

To decrease the effects of frequency dispersion on the interpolated wavefield the interpolation weights can be made frequency dependent by recognizing that the effective value for the horizontal wavenumber \widehat{k}_m , which enters in the computation of $\Delta k_z / \Delta s$ in equation (26), is an approximation of the exact wavenumber k_m . When the second derivatives are computed using the first order approximation T_b in equation (18), \widehat{k}_m is given as a function of k_m by the following expression,

$$\widehat{k}_m = \frac{2}{\Delta x} \sin \frac{k_m \Delta x}{2}. \quad (27)$$

It is important to notice that the stability analysis developed in the previous section strictly applies only to the simple FFD correction, not to its combination with an interpolation scheme like FFDPI. In theory, instability can still develop when using FFDPI, as it does for PSPI (Margrave and Ferguson, 1999). However, in my tests I have not encountered it yet!

FFDPI error analysis

Figure 9 shows the effect of interpolation on the phase curves after interpolation. The phase function was interpolated after the application of the FFD correction starting from a reference velocity lower than the medium velocity and a reference velocity higher than the medium velocity. As in Figure 2, the medium velocity v is equal to 2 km/s, and two reference velocities are assumed: one 10% lower than the medium velocity (1.8 km/s), the other one 10%

higher than the medium velocity (2.2 km/s). The interpolation weights were computed using equation (26) at an angle of 60 degrees.

Figure 10 shows the phase error measured at 50 degrees as a function of the medium velocity v as it spans the range between the lower reference velocity (1.8 km/s) and the higher reference velocity (2.2 km/s). It compares the phase errors at three different frequencies (0 Hz, 30 Hz and 60 Hz) when the interpolation weights are computed assuming an exact second derivative operator (i.e. weights are constant with frequency) and the first order approximation of the second derivative operator (i.e. weights vary with frequency). The absolute value of the maximum error suffered when frequency-dependent weights are used ($\approx 0.1\%$ at 60 Hz and $v \approx 1.95$ km/s) is very small, and almost a factor of 8 smaller than when frequency-independent weights are used.

Figures 11 and 12 show the impulse responses corresponding to the phase plots shown in Figures 9 and 10 and should be compared with the impulse responses shown in Figures 3–5. The impulse response shown in Figure 11 was computed using frequency-independent interpolation weights. While it is much closer to the exact impulse response (Figure 3) than either of the impulse responses obtained using a simple FFD correction (Figures 4 and 5), it shows some frequency dispersion. The higher frequencies are imaged inside the semicircle. The frequency dispersion is greatly reduced when the frequency-dependent interpolation weights are used, as demonstrated in Figure 12, and predicted by the plots in Figure 10.

A wavefield interpolation scheme similar to the one described above could be used in conjunction with the split-step correction instead of the FFD correction. With a fixed number of reference velocities, using split step instead of FFD would reduce the computational cost. It is thus useful to compare the phase errors of the two competing methods, to analyze the improvement in accuracy gained using the more accurate but more expensive correction method. Figure 13 compares the phase errors measured at 50 degrees as a function of the medium velocity v as it spans the range between the lower reference velocity (1.8 km/s) and the higher reference velocity (2.2 km/s). The maximum error suffered by the method based on the split-step correction is about 8 times larger than the error suffered when the FFD correction is applied.

Azimuthal anisotropy

A recurring problem that hampers the application of implicit finite-difference methods to 3-D wave extrapolation is the azimuthal anisotropy associated with splitting (Jakubowicz and Levin, 1983). Of course, this problem affects also the FFD correction applied by splitting (1998). One potentially attractive way of solving this problem is using helical boundary conditions, as Rickett discusses in an article in this report (2000). However, I will show that for the FFDPI algorithm the azimuthal anisotropy is greatly reduced without recurring to sophisticated linear solvers.

Figure 14 compares relative phase errors as a function of the azimuth measured for a propagation angle of 53 degrees. As in the previous figures, the medium velocity v is equal to 2 km/s, and two reference velocities are assumed: one 10% lower than the medium velocity (1.8 km/s),

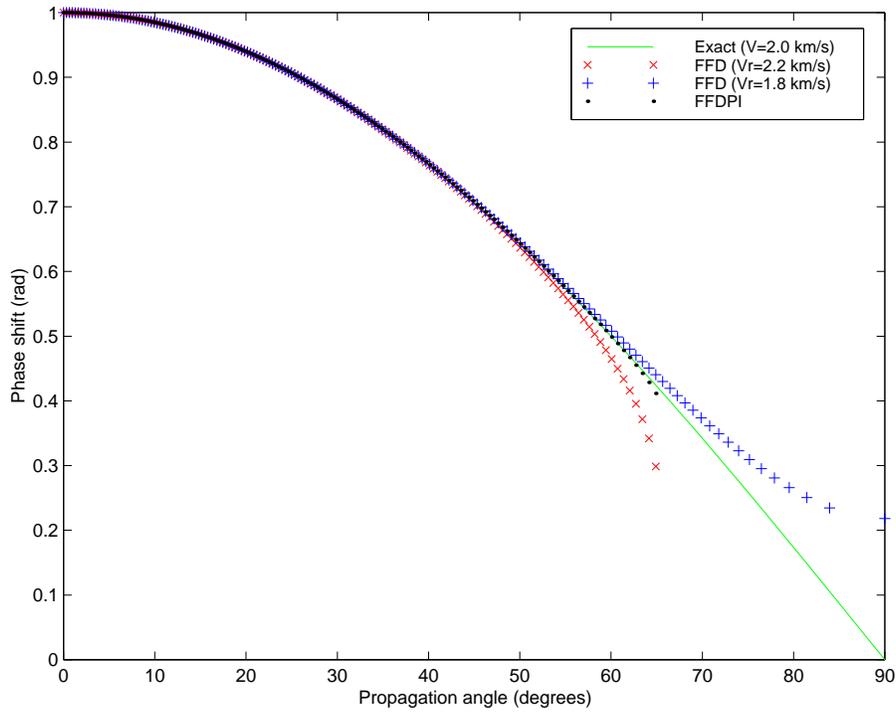


Figure 9: Phase curves for: (-) Exact with $v=2$ km/s, (x) FFD with $v=2.2$ km/s, (+) FFD with $v=1.8$ km/s, (.) FFDPI `biondo1-interp` [CR]

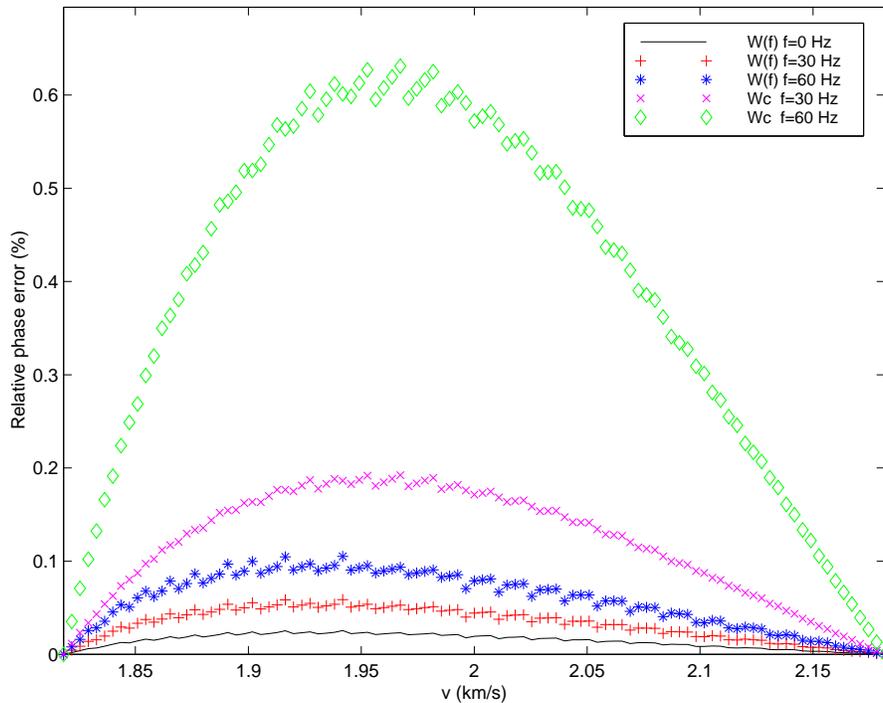


Figure 10: Relative phase errors curves for: (-) Zero frequency, (+) Frequency-dependent weights at 30 Hz. (+) Frequency-dependent weights at 60 Hz. (x) Frequency-independent weights at 30 Hz. (\diamond) Frequency-independent weights at 60 Hz. `biondo1-errfreq` [CR]

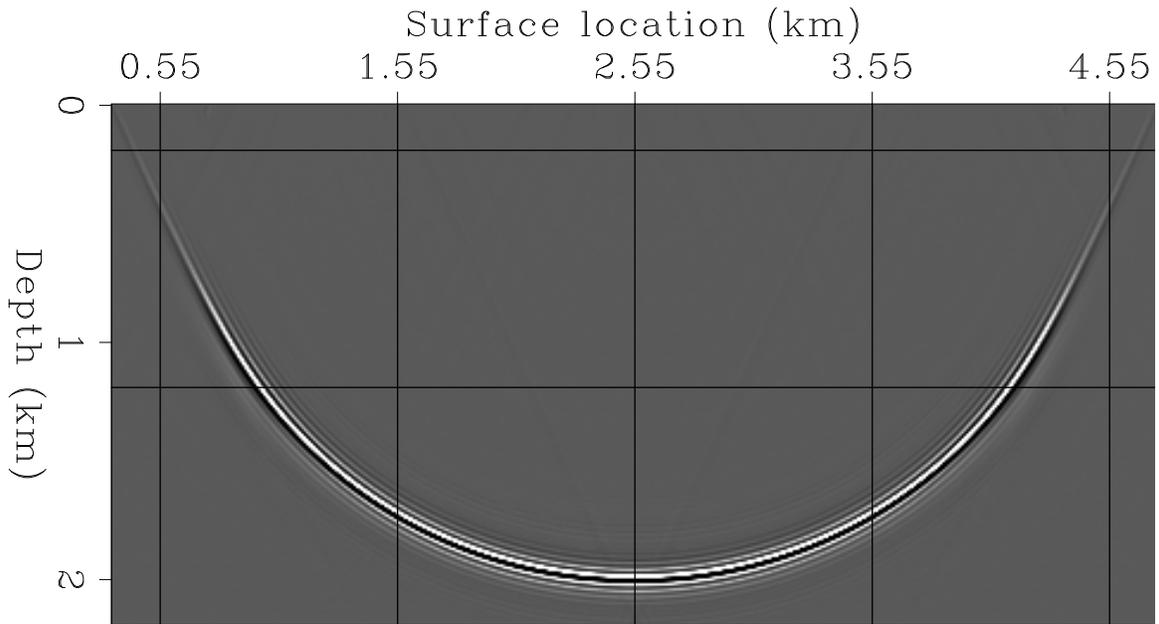


Figure 11: Impulse response after interpolation with frequency-independent weights.
`biondo1-Line-rist-interp-ndisp` [ER]

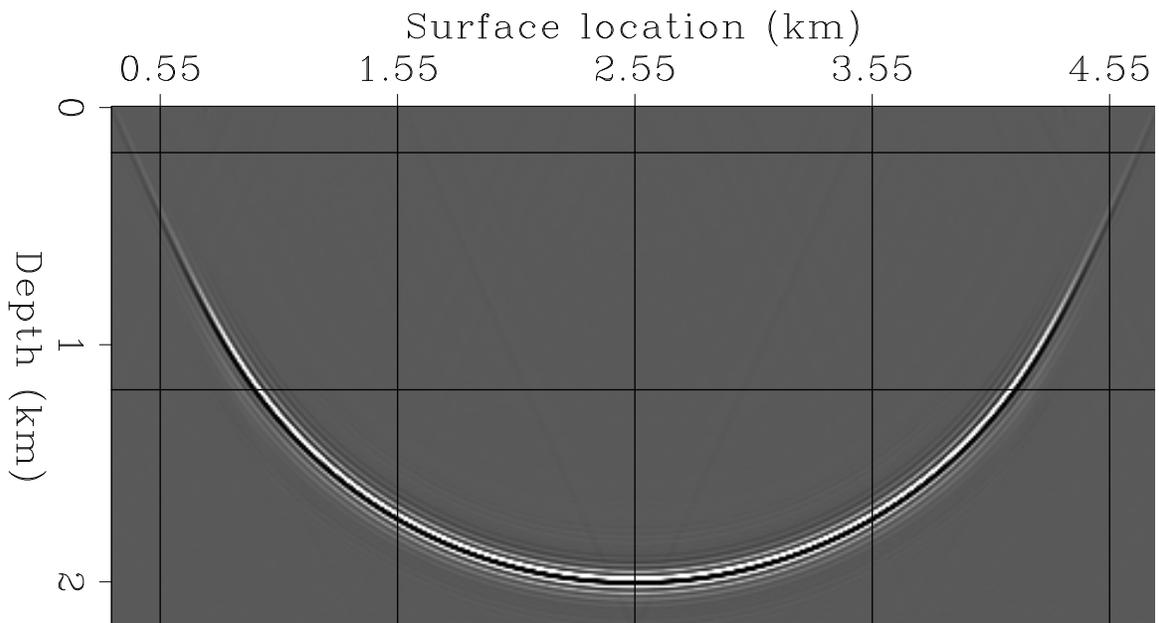


Figure 12: Impulse response after interpolation with frequency-dependent weights.
`biondo1-Line-rist-interp` [ER]

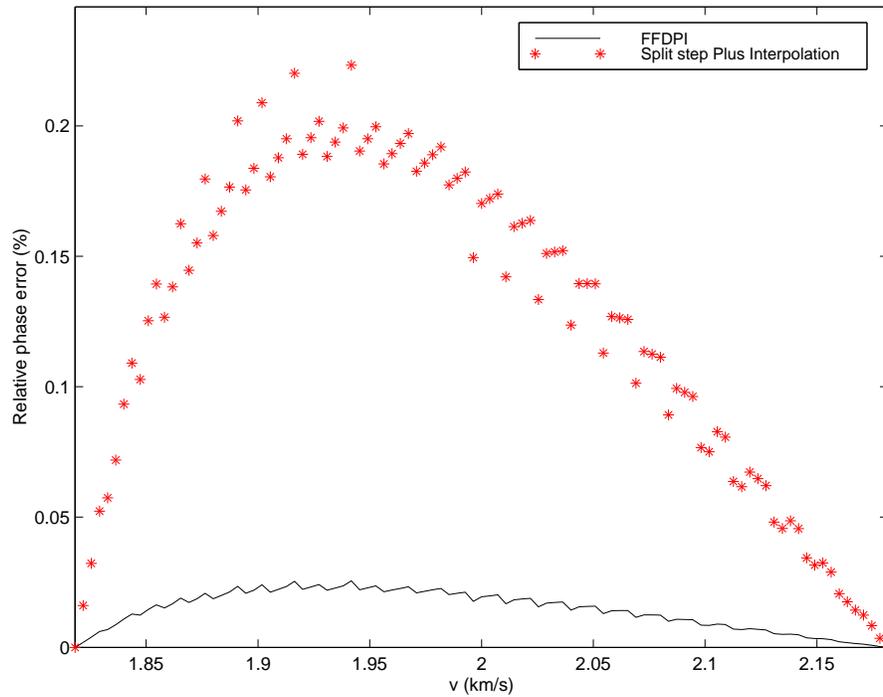


Figure 13: Relative phase errors curves after interpolation for: (-) FFD correction, (+) Split-step correction. `biondo1-errint` [CR]

the other one 10% higher than the medium velocity (2.2 km/s). The frequency-dependent interpolation weights were computed to zero the phase error along the inline/crossline directions for the same propagation angle. The plots show the phase errors at two frequencies (0 Hz and 60 Hz) for the FFDPI algorithm, the FFD correction starting from the lower reference velocity, and the FFD correction starting from the higher reference velocity. Notice that for both the simple FFD correction cases the azimuthal anisotropy decreases as the frequency increases, though the average phase error increase as well. But the crucial, and useful, feature of the phase errors function for the FFD corrections is that the azimuthal variations are in the opposite direction when the differences between the reference velocity and medium velocity have opposite sign. Therefore, the phase error of the interpolation method is much lower than the error of either of the simple FFD corrections. At higher frequencies (60 Hz) the impulse response of FFDPI is almost perfectly isotropic.

The theoretical analysis is confirmed by the characteristics of the impulse responses. Figure 15 shows the depth slice of three impulse responses superimposed to each other. The outermost circular event corresponds to the FFD correction starting from a reference velocity of 2.2 km/s. The middle event corresponds to the exact impulse response with the medium velocity of 2 km/s. The innermost event corresponds to the FFD corrections starting from a reference velocity of 1.8 km/s. As predicted by the plots shown in Figure 14, the azimuthal anisotropy is frequency dependent and the frequency dispersion is smaller for azimuths oriented at 45 degrees with respect to the coordinate axes.

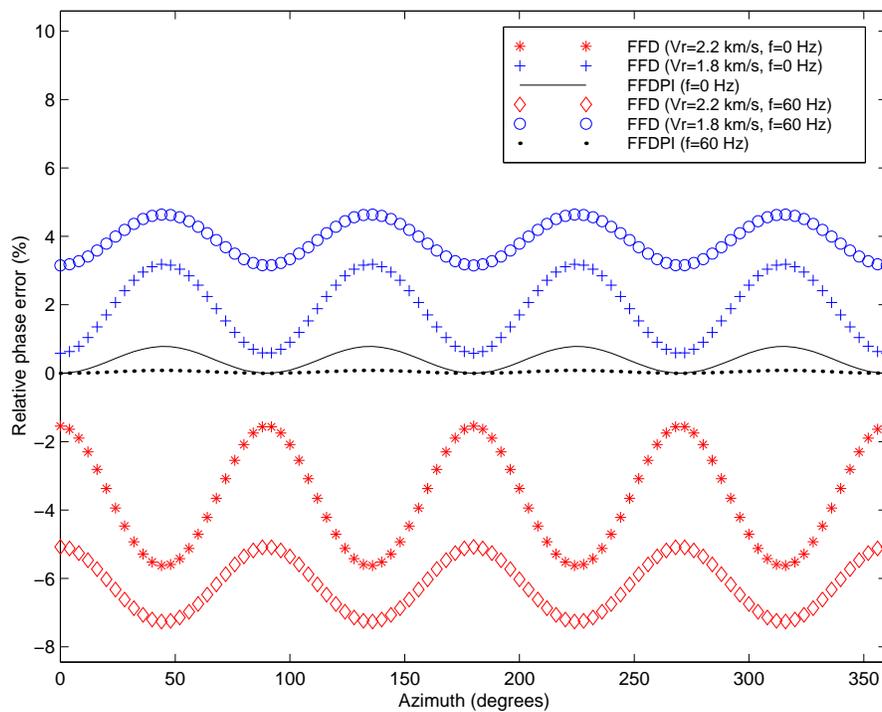


Figure 14: Relative phase errors curves for: (-) FFDPI at 0 Hz, (.) FFDPI at 60 Hz, (+) FFD with $v=2.2$ km/s at 30 Hz, (\diamond) FFD with $v=2.2$ km/s at 60 Hz, (+) FFD with $v=1.8$ km/s at 30 Hz, (\circ) FFD with $v=1.8$ km/s at 60 Hz. `biondo1-azim` [CR]

Figure 16 is the merge of two impulse responses along the inline direction. For negative values of the in-line coordinate the plot shows the depths slice for the exact impulse response. For positive values of the in-line coordinate the plot shows the depths slice for the impulse response obtained by FFDPI. It is evident that the result of the interpolation scheme is much less affected by azimuthal anisotropy and frequency dispersion than the results of the two simple FFD correction showed in Figure 15.

ZERO-OFFSET MIGRATION OF THE SEG-EAGE SALT DATA SET

To test both the stability and the accuracy of the FFDPI algorithm, I migrated the zero-offset data from the SEG/EAGE salt data set (Aminzadeh et al., 1996). The data set is a good test for the stability of the FFDPI algorithm because the velocity model has sharp discontinuities caused by the salt body. Furthermore, because of a low-velocity region intended to model subsalt overpressure, several depth slices have a wide range of velocities. Figure 17 shows one of these depth slices. In the plot the salt velocity is clipped, thus the scale-bar on the side represents the range of velocities within the sediments. There is almost a factor of two between the slow velocity sediments in the ‘overpressure zone’ in the middle, and the faster sediments at the edges.

The migration algorithm does not need to handle accurately lateral velocity variations to

Figure 15: Depth slices through impulse responses: 1) innermost event corresponds to the FFD corrections starting from a reference velocity of 1.8 km/s, 2) middle event corresponds to the exact impulse response with the medium velocity of 2 km/s, 3) outermost event corresponds to the FFD corrections starting from a reference velocity of 1.8 km/s. [biondo1-Aniso-rist](#) [CR]

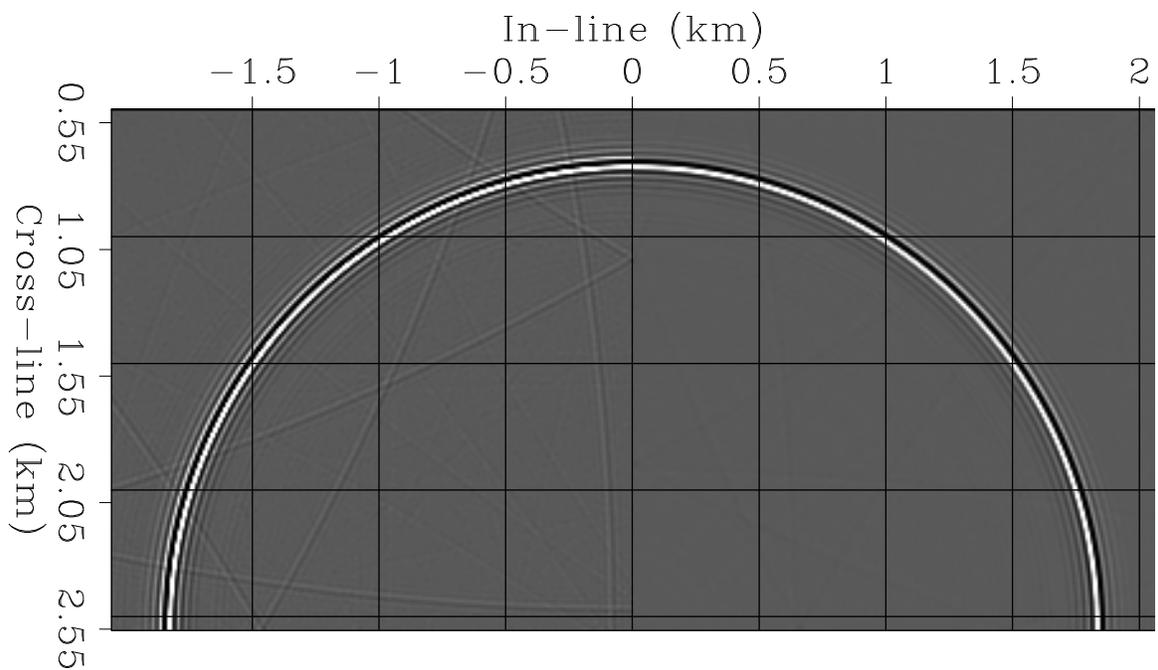
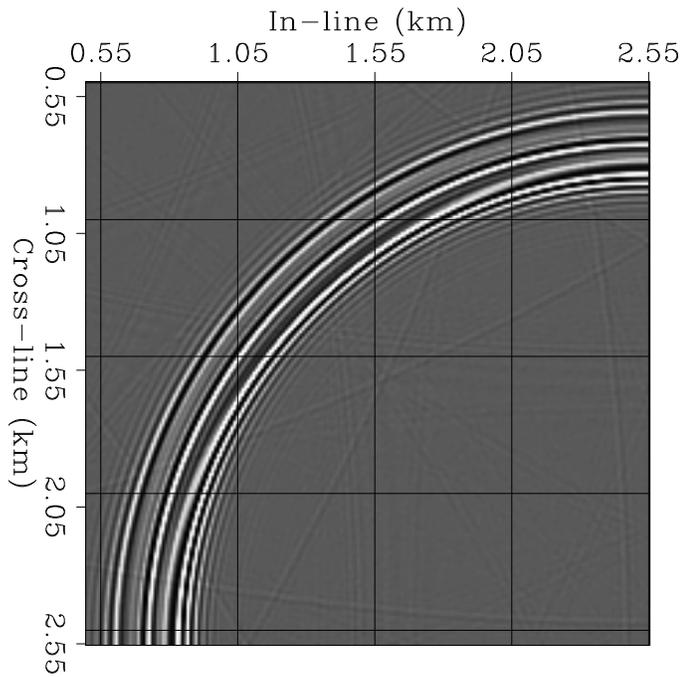


Figure 16: Depth slices through impulse responses: 1) left half corresponds to the exact impulse response with the medium velocity of 2 km/s, 2) right half corresponds to the interpolation method results. [biondo1-Iso-rist](#) [CR]

image the reflectors above the salt. While the reflectors below the salt cannot be imaged by simple zero-offset migration. A deep fault (between depths of 2 km and 3 km) and away from the salt body is one of the few reflectors that is well suited to test the accuracy of a zero-offset migration. In previous reports I noted that six reference velocities are needed to image this particular fault by common-azimuth migration employing a split-step correction (Biondi, 1999b,a). Figure 18 shows an in-line section of the migrated cube that cuts across the fault of interest. It was obtained using the FFDPI algorithm. Four reference velocities were used at each depth step. Figures 19 and 20 compare the zooms around the fault of interest. The light (yellow) line superimposed onto both plots represents the correct fault position, as picked from the velocity model. Figure 19 shows the results obtained by using the interpolation algorithm described above, but in conjunction with the split-step correction instead of the stable FFD correction. The fault is undermigrated and thus it is imaged too shallow, and the sediment terminations are not well focused. The fault is better positioned and the image is better focused when the FFDPI algorithm is used to migrate the data (Figure 20).

CONCLUSIONS

The combination of Fourier methods' accuracy for wide-angle propagation with implicit finite differences' flexibility for modeling lateral velocity variations yields accurate and efficient downward-propagation methods. The FFD corrections is the most attractive among several methods proposed in the literature to correct constant-velocity phase shift for lateral velocity variations. However, the correction operator originally presented by Ristow and Ruhl (1994) can be unstable in presence of sharp discontinuities in the velocity function. In this paper I present and successfully test a stable version of the FFD correction.

Using the stable FFD correction as a building block, I derive an accurate and stable wide-angle migration (Fourier-Finite Difference Plus Interpolation). The FFDPI algorithm is based on the interpolation of two wavefields corrected using the FFD method with opposite signs of the velocity perturbations. This interpolation step compensates for both the azimuthal anisotropy and the frequency-dispersion of the simple FFD corrections. Therefore the FFDPI algorithm achieves high accuracy, as demonstrated by the migration example of the SEG-EAGE salt data set. The accuracy and the cost of FFDPI algorithm can be easily controlled by setting the number of reference velocities.

REFERENCES

- Aminzadeh, F., Burkhard, N., Long, J., Kunz, T., and Duclos, P., 1996, Three dimensional SEG/EAGE models - an update: *The Leading Edge*, **2**, 131–134.
- Biondi, B., and Palacharla, G., 1996, 3-D prestack migration of common-azimuth data: *Geophysics*, **61**, no. 6, 1822–1832.
- Biondi, B., 1999a, Offset plane waves vs. common-azimuth migration for sub-salt imaging: *SEP-102*, 15–34.

Figure 17: Depth slice of the velocity model at $z=2.1$ km. The salt velocity was clipped, thus the scale bar on the side shows the range of velocities in the sediments.

biondo1-Vel_spk_z2100 [ER]

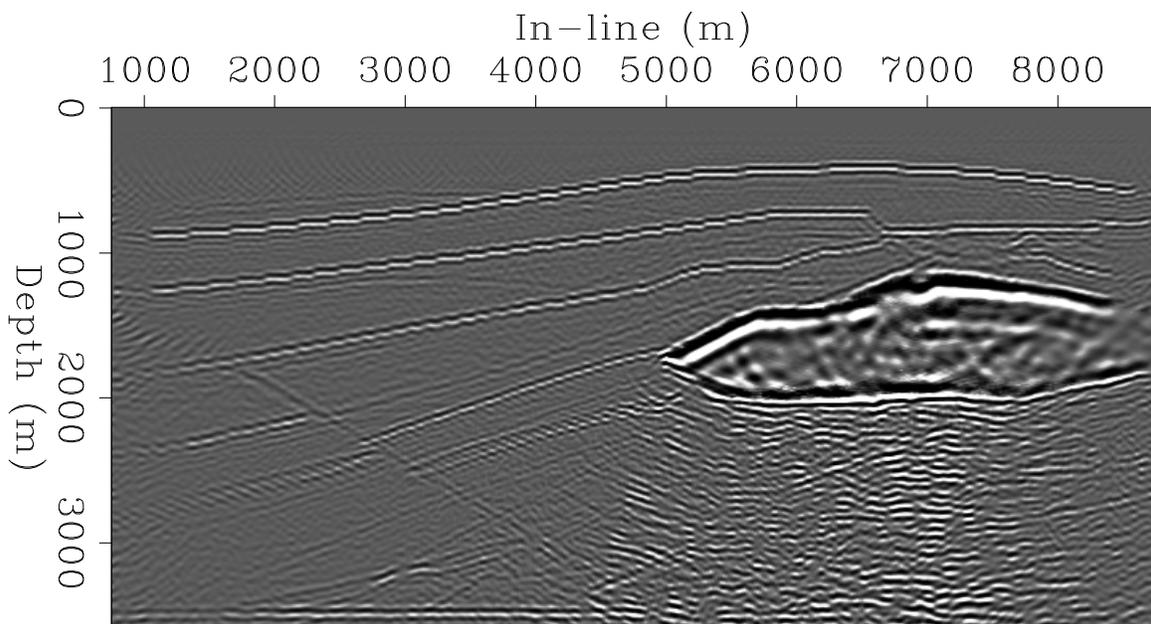
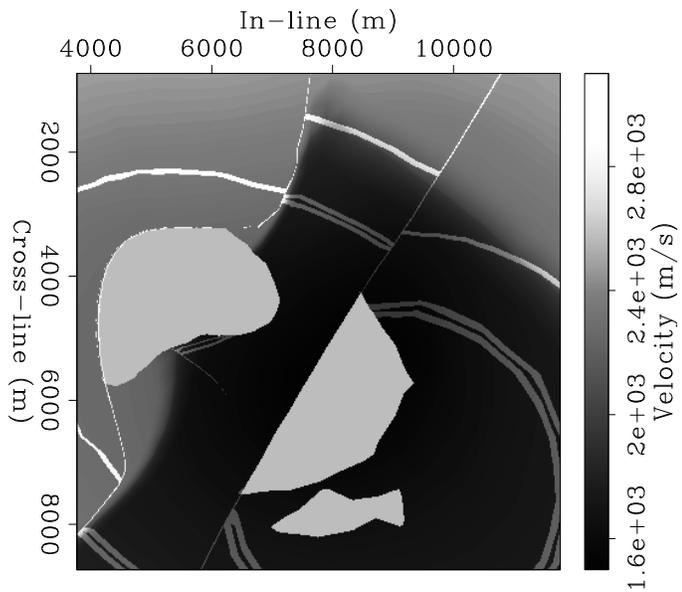


Figure 18: In-line section of the migrated cube obtained by use of FFDPI.

biondo1-Line-rist2-y9820 [CR]

Figure 19: Window of the same in-line section shown in Figure 18, but obtained by use of split step to correct for lateral velocity variations. The light (yellow) line superimposed onto the plot represents the correct fault position, as picked from the velocity model. Notice the misplacement of the fault. `biondo1-salt-spl` [NR]

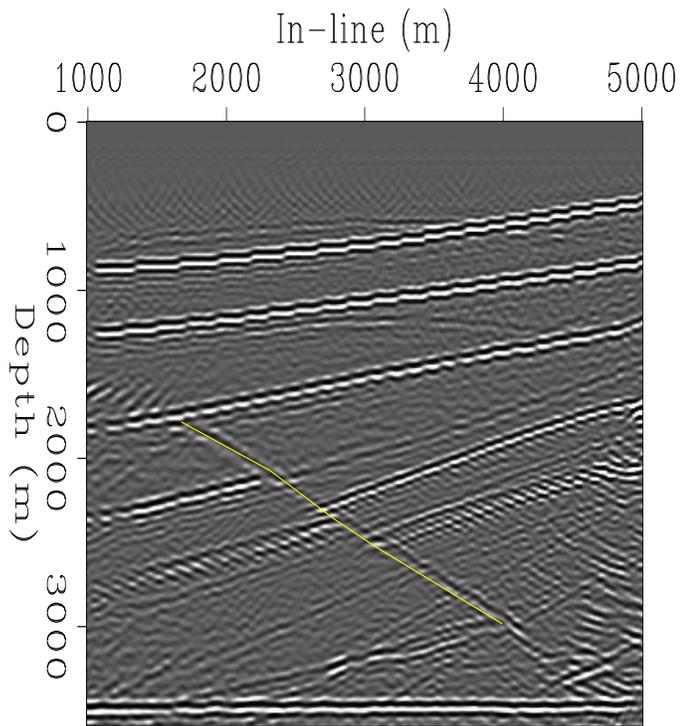
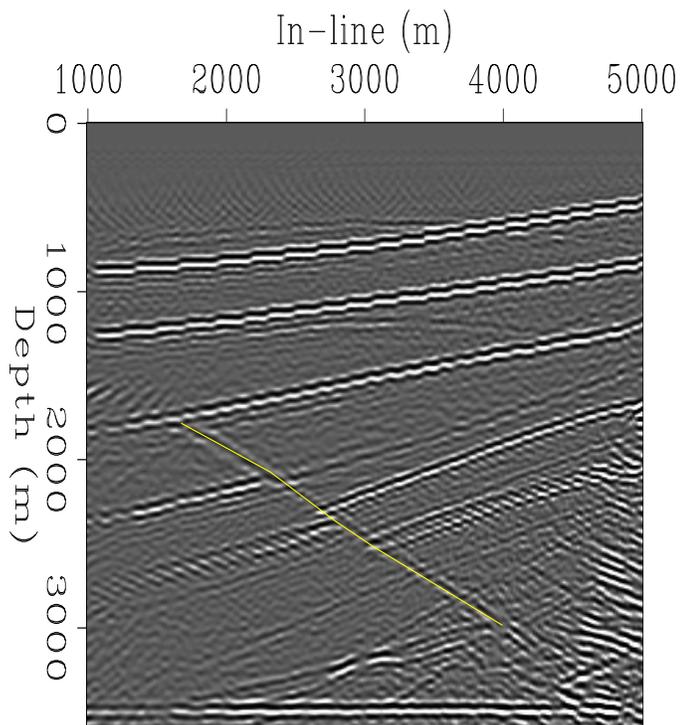


Figure 20: Window of the same in-line section shown in Figure 18 and obtained by use of FFDPI. The light (yellow) line superimposed onto the plot represents the correct fault position, as picked from the velocity model. Notice the better placement of the fault. `biondo1-salt-rist` [NR]



- Biondi, B., 1999b, Subsalt imaging by common-azimuth migration: SEP-**100**, 113–124.
- Brown, D., 1979, Muir's rules for matrices: Another look at stability: SEP-**20**, 125–142.
- Claerbout, J. F., 1985, *Imaging the Earth's Interior*: Blackwell Scientific Publications.
- Clayton, R. W., and Engquist, B., 1977, Absorbing boundary conditions: SEP-**11**, 116–132.
- Cockshott, I., and Jakubowicz, H., 1998, 3-D migration using the Fourier finite-difference method: 66th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 435–438.
- Gazdag, J., and Sguazzero, P., 1984, Migration of Seismic Data by Phase-Shift Plus Interpolation: *Geophysics*, **49**, no. 2, 124–131.
- Gazdag, J., 1978, Wave equation migration with the phase-shift method: *Geophysics*, **43**, no. 10, 1342–1351.
- Godfrey, R. J., Muir, F., and Claerbout, J. F., 1979, Stable extrapolation: SEP-**16**, 83–87.
- Huang, L.-J., Fehler, M. C., and Wu, R. S., 1999, Extended local Born Fourier migration method: *Geophysics*, **64**, no. 5, 1524–1534.
- Jacobs, B., and Muir, F., 1981, Convergence of the continued fraction for the square root function: SEP-**26**, 183–196.
- Jakubowicz, H., and Levin, S., 1983, A simple exact method of three-dimensional migration – Theory: *Geophys. Prosp.*, **31**, no. 1, 34–56.
- Jin, S., Wu, R. S., and Peng, C., 1998, Prestack depth migration using a hybrid pseudo-screen propagator: 68th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1819–1822.
- Le Rousseau, J. H., and de Hoop, M. V., 1998, Modeling and imaging with the generalized screen algorithm: 68th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1937–1940.
- Margrave, G., and Ferguson, R., 1999, An explicit, symmetric wavefield extrapolator for depth migration: 69th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1461–1464.
- Mosher, C. C., Foster, D. J., and Hassanzadeh, S., 1997, Common angle imaging with offset plane waves: 67th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1379–1382.
- Rickett, J., 2000, Efficient 3-D wavefield extrapolation with Fourier finite-differences and helical boundary conditions: SEP-**103**, 103–109.
- Ristow, D., and Ruhl, T., 1994, Fourier finite-difference migration: *Geophysics*, **59**, no. 12, 1882–1893.

Rothman, D. H., and Thorson, J. R., 1982, An analysis of stable extrapolation operators with absorbing boundaries: *SEP-32*, 105–116.

Stoffa, P. L., Fokkema, J. T., de Luna Freire, R. M., and Kessinger, W. P., 1990, Split-step Fourier migration: *Geophysics*, **55**, no. 4, 410–421.

Efficient 3-D wavefield extrapolation with Fourier finite-differences and helical boundary conditions

James Rickett¹

ABSTRACT

Fourier finite-difference (FFD) migration combines the complementary advantages of the phase-shift and finite-difference migration methods. However, as with other implicit finite-difference algorithms, direct application to 3-D problems is prohibitively expensive. Rather than making the simple $x - y$ splitting approximation that leads to extensive azimuthal operator anisotropy, I demonstrate an alternative approximation that retains azimuthal isotropy without the need for additional correction terms. Helical boundary conditions allow the critical 2-D inverse-filtering step to be recast as 1-D inverse-filtering. A spectral factorization algorithm can then factor this 1-D filter into a (minimum-phase) causal component and a (maximum-phase) anti-causal component. This factorization provides an LU decomposition of the matrix, which can then be inverted directly by back-substitution. The cost of this approximate inversion remains $O(N)$ where N is the size of the matrix.

INTRODUCTION

Within the exploration industry, geophysicists are realizing the inherent limitations of Kirchhoff methods when it comes to accurately modeling the effects of finite-frequency wave propagation. This is fueling interest in “wave-equation” migration algorithms, such as those based on wavefield extrapolation, that do accurately model finite-frequency effects.

As with all migration algorithms, there is a tradeoff amongst extrapolators: cost versus accuracy. For wavefield extrapolators, however, the tradeoff goes three ways: accuracy at steep dips versus the ability to accurately handle lateral velocity variations versus cost again. Fourier finite-difference migration (Ristow and Ruhl, 1994) strikes an effective balance between the accuracy priorities, combining the steep dip accuracy of phase-shift migration in media with weak lateral velocity contrasts, and the ability to handle lateral variations with finite-difference.

Unfortunately, as with other implicit finite-difference, the cost does not scale well for three-dimensional problems without additional approximations that often expensive and may compromise accuracy. In an earlier paper, Rickett et al. (1998) solved the costly matrix inversion for implicit extrapolation with the 45° equation with an approximate LU decomposition

¹email: james@sep.stanford.edu

based on the helical transform (Claerbout, 1998b). In this paper, the same approach allows me to extrapolate with a more accurate operator.

THEORY

Fourier finite-difference migration

Three-dimensional FFD extrapolation is based on the equation (Ristow and Ruhl, 1994),

$$\frac{\partial P}{\partial z} = i \left[\sqrt{\frac{\omega^2}{c^2} + \nabla_{x,y}^2} + \left(\frac{\omega}{v} - \frac{\omega}{c} \right) + \frac{\omega}{v} \left(1 - \frac{c}{v} \right) \frac{\frac{v^2}{\omega^2} \nabla_{x,y}^2}{a + b \frac{v^2}{\omega^2} \nabla_{x,y}^2} \right] P, \quad (1)$$

where $v = v(x, y, z)$ is the medium velocity, c is a reference velocity ($c \leq v$), and a and b are coefficients subject to optimization. The first term describes a simple Gazdag phase-shift that must be applied in the (ω, \mathbf{k}) domain; the second term describes the first-order split-step correction (Stoffa et al., 1990), applied in (ω, \mathbf{x}) ; and the third term describes an additional correction that can be applied as an implicit finite-difference operator (Claerbout, 1985), also applied in (ω, \mathbf{x}) .

In areas with strong lateral velocity variations ($c/v \approx 0$), FFD reduces to a finite-difference migration, while in areas of weak lateral velocity variations ($c/v \approx 1$), FFD retains the steep-dip accuracy advantages of phase-shift migration. As a full-wave migration method, FFD also correctly handles finite-frequency effects.

For constant lateral velocity, the finite-difference term in equation (1) can be rewritten as the following matrix equation,

$$(\mathbf{I} + \alpha_1 \mathbf{D}) \mathbf{q}_{z+1} = (\mathbf{I} + \alpha_2 \mathbf{D}) \mathbf{q}_z \quad (2)$$

$$\mathbf{A}_1 \mathbf{q}_{z+1} = \mathbf{A}_2 \mathbf{q}_z \quad (3)$$

where \mathbf{D} is a finite-difference representation of the x, y -plane Laplacian, $\nabla_{x,y}^2$, and \mathbf{q}_z and \mathbf{q}_{z+1} represent the diffraction wavefield at depths z and $z + 1$ respectively. Scaling coefficients, α_1 and α_2 , are complex and depend both on the ratio, ω/v , and the ratio c/v .

The right-hand-side of equation (3) is known. The challenge is to find the vector \mathbf{q}_{z+1} by inverting the matrix, \mathbf{A}_1 . For 2-D problems, only a tridiagonal matrix must be inverted; whereas, for 3-D problems the matrix becomes blocked tridiagonal. For most applications, direct inversion of such a matrix is prohibitively expensive, and so approximations are required for the algorithm to remain cost competitive with other migration methods.

A partial solution is to split the operator to act sequentially along the x and y axes. Unfortunately this leads to extensive azimuthal operator anisotropy, and necessitates expensive additional phase correction operators.

Helical factorization

The blocked-tridiagonal matrix of the 3-D extrapolation, \mathbf{A}_1 , represents a two-dimensional convolution operator. Following Rickett et al.'s (1998) approach to factoring the 45° equation, I apply helical boundary conditions (Claerbout, 1998b) to simplify the structure of the matrix, reducing the 2-D convolution to an equivalent problem in one dimension.

For example, helical boundary conditions allow a two-dimensional 5-point Laplacian filter to be expressed as an equivalent one-dimensional filter of length $2N_x + 1$ as follows

$$d = \begin{bmatrix} & & 1 & & \\ & 1 & -4 & 1 & \\ & & & & \\ & & & & 1 \\ & & & & \end{bmatrix} \xrightarrow{\text{helical boundary conditions}} (1, 0, \dots, 0, 1, -4, 1, 0, \dots, 0, 1).$$

The operator, \mathbf{D} , in equation (2) could represent convolution with this filter; however, I use a more accurate, but equivalent, 9-point filter.

Unfortunately, the complex scale-factor, α_1 , means \mathbf{A}_1 is symmetric, but not Hermitian, so the filter, a_1 , is not an autocorrelation function, and standard spectral factorization algorithms will fail. Fortunately, however, the Kolmogoroff method can be extended to factor any cross-spectrum into a pair of minimum phase wavelets and a delay (Claerbout, 1998a).

With this algorithm, the 1-D convolution filter of length $2N_x + 1$ can be factored into a pair of (minimum-phase) causal and (maximum-phase) anti-causal filters, each of length $N_x + 1$. Fortunately, filter coefficients drop away rapidly from either end, and in practice, small-valued coefficients can be safely discarded.

By reconstituting the matrices representing convolution with these filters, I obtain an approximate LU decomposition of the original matrix. The lower and upper-triangular factors can then be inverted efficiently by recursive back-substitution.

While we have only described the factorization for $v(z)$ velocity models, the method can also be extended to handle lateral variations in velocity. For every value of ω/v and c/v , we precompute the factors of the 1-D helical filters, a_1 and a_2 . Filter coefficients are stored in a look-up table. We then extrapolate the wavefield by non-stationary convolution, followed by non-stationary polynomial division. The convolution is with the spatially variable filter pair corresponding to a_2 . The polynomial division is with the filter pair corresponding to a_1 . The non-stationary polynomial division is exactly analogous to time-varying deconvolution, since the helical boundary conditions have converted the two-dimensional system to one-dimension.

EXAMPLES

Figure 1 compares depth-slices through impulse responses of FFD migration (with $c/v = 0.8$) for the splitting approximation, (a), and the helical factorization, (b). The azimuthal anisotropy is noticeably reduced with the helical factorization.

Figure 2 shows extracts from a three-dimensional FFD depth migration of a zero-offset subset from the SEG/EAGE salt dome dataset. This rugose lateral velocity model initially

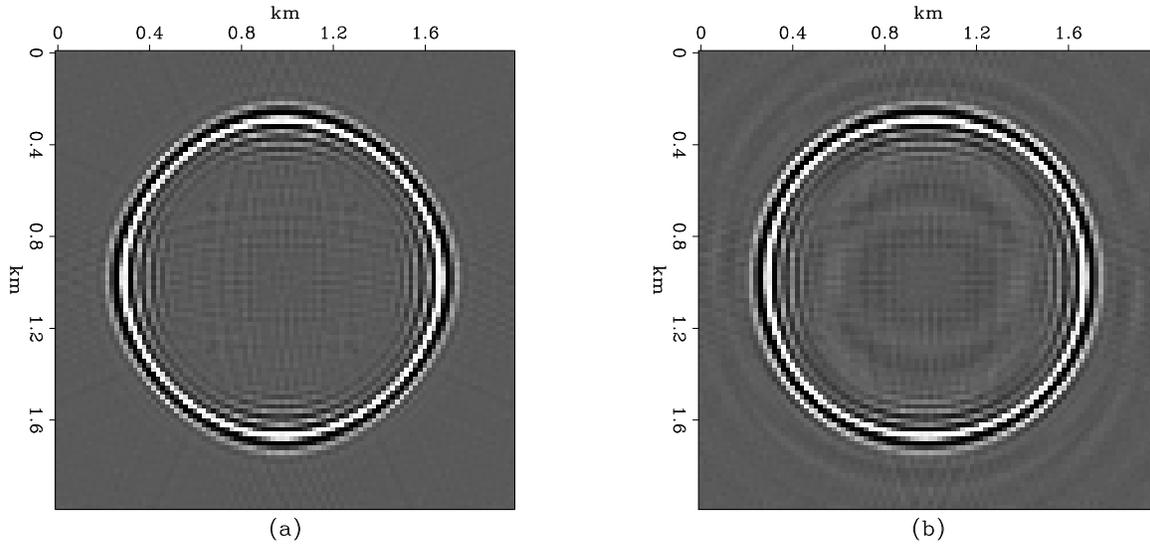


Figure 1: Depth-slices of centered impulse response corresponding to a dip of 45° for $c/v = 0.8$. Panel (a) shows the result of employing an $x - y$ splitting approximation, and panel (b) shows the result of the helical factorization. Note the azimuthally isotropic nature of panel (b).
james1-timeslices [CR]

caused mild stability problems for the FFD migration, and I had to smooth the velocity model to produce the results shown in Figure 2. Biondi (2000) presents an unconditionally stable formulation of the FFD algorithm; however, that formulation does not easily fit with the approximate helical factorization discussed here.

CONCLUSIONS

Helical boundary conditions allow the critical 2-D inverse-filtering step in FFD migration to be recast as 1-D inverse-filtering. A spectral factorization algorithm can then factor this 1-D filter into a (minimum-phase) causal component and a (maximum-phase) anti-causal component. This factorization provides an LU decomposition of the matrix, which can then be inverted directly by back-substitution. The cost of this approximate inversion remains $O(N)$ where N is the size of the matrix.

I demonstrate this alternative factorization retains azimuthal isotropy without the need for additional correction terms, and apply the migration algorithm to the 3-D SEG/EAGE salt dome synthetic dataset.

REFERENCES

Biondi, B., 2000, Stable wide-angle Fourier-finite difference downward extrapolation of 3-D wavefields: SEP-103, 79–103.

- Claerbout, J. F., 1985, *Imaging the earth's interior*: Blackwell Scientific Publications.
- Claerbout, J., 1998a, Factorization of cross spectra: SEP-**97**, 337–342.
- Claerbout, J., 1998b, Multidimensional recursive filters via a helix: *Geophysics*, **63**, 1532–1541.
- Rickett, J., Claerbout, J., and Fomel, S., 1998, Implicit 3-D depth migration by wavefield extrapolation with helical boundary conditions: 68th Ann. Internat. Meeting, Soc. Expl. Geophys., Expanded Abstracts, 1124–1127.
- Ristow, D., and Ruhl, T., 1994, Fourier finite-difference migration: *Geophysics*, **59**, no. 12, 1882–1893.
- Stoffa, P. L., Fokkema, J. T., de Luna Freire, R. M., and Kessinger, W. P., 1990, Split-step Fourier migration: *Geophysics*, **55**, no. 4, 410–421.

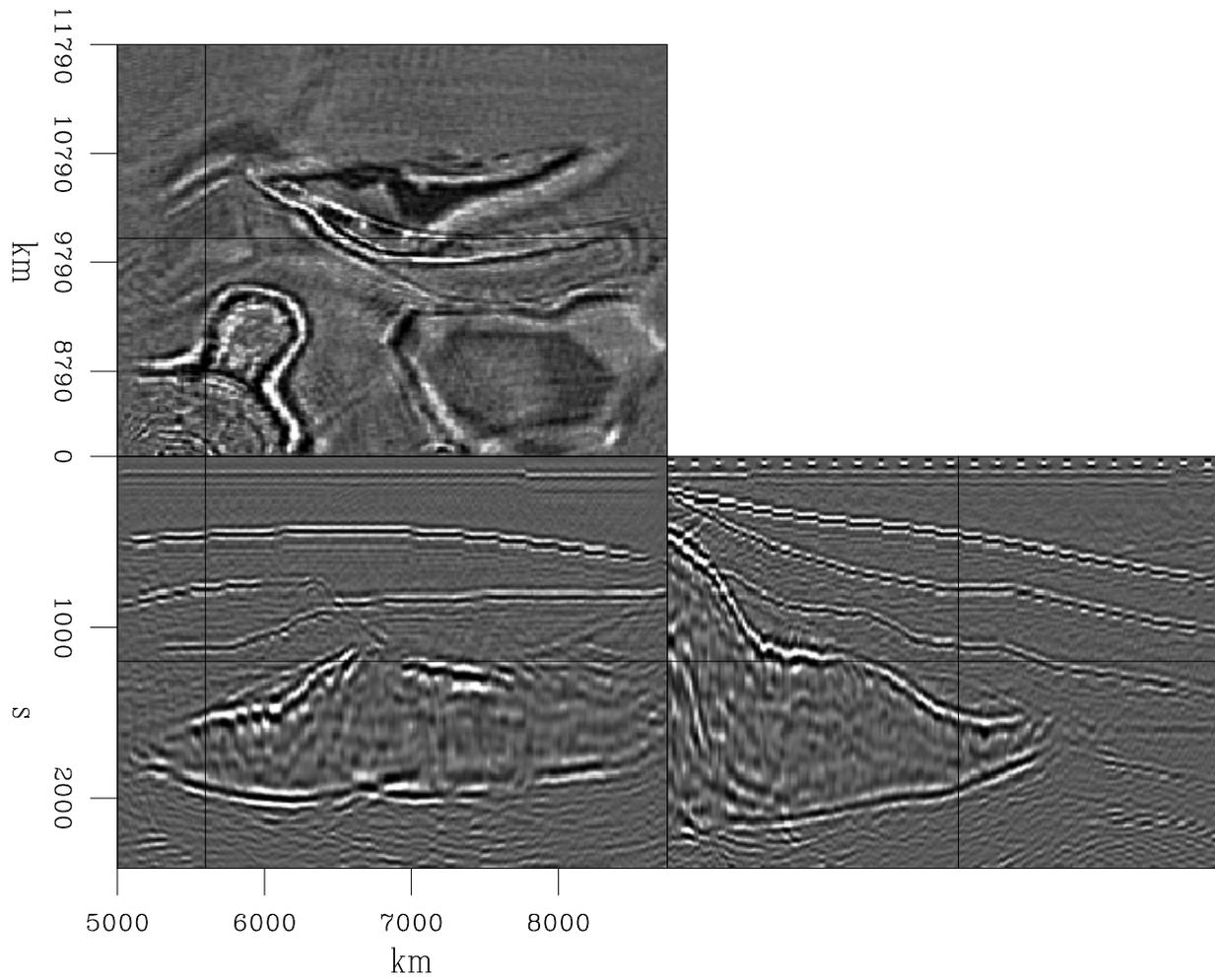


Figure 2: Migration of a three-dimensional zero-offset subset from the SEG/EAGE salt dome dataset by Fourier finite-differences with helical boundary conditions. `james1-cubeplot` [CR]

Seismic image regularization in the reflection angle domain

Marie L. Prucha, Robert G. Clapp, and Biondo Biondi¹

ABSTRACT

We explore the use of preconditioned inversion in the reflection angle domain rather than migration to improve imaging in complex media. We use a wave-equation method to create reflection angle domain common image gathers and we apply steering filter preconditioning to smooth along the reflection angles. This improves the common image gathers. The improved common image gathers are more continuous than common image gathers obtained by migration alone. Additionally, some multiple energy is attenuated.

INTRODUCTION

Current migration methods are done mostly in the offset domain and the shot domain, which are prone to multipathing (ten Kroode et al., 1999) in complex areas. A fairly new method that is being used to image complex areas is migration in the reflection angle domain (RAD). This can be done both by Kirchhoff methods (Xu et al., 1998) and wave-equation methods (Prucha et al., 1999). The RAD avoids the problem of multipathing, and therefore contains fewer artifacts than the more commonly used domains. Unfortunately, even with fewer multipathing artifacts, in complex areas migration may not be enough. Since we are interested in complex areas, we can reformulate our imaging problem as an inversion problem (Chemingui, 1999).

Although imaging by inversion can give better results than migration, an inversion problem can be unstable (Claerbout, 1991). A trick used to constrain inversion problems to a reasonable result is regularization (Harlan, 1986; Fomel, 1997). Theory states that for a particular point in the subsurface, the reflectivity as a function of reflection angle should vary smoothly (Richter, 1941). Therefore, the obvious choice for a regularization operator in the reflection angle domain is one that smooths along the reflection angles. To speed the convergence, we can reformulate the regularization problem as a preconditioned problem (Fomel et al., 1997). We intend to show that applying this method in the reflection angle domain will improve the common image gathers (CIGs), making the events more continuous, reducing artifacts, and attenuating multiples.

In this paper, we will first explain how to image in the reflection angle domain and how to apply regularization and preconditioning. Then we will show the results of applying regularization to a RAD inversion problem on two different synthetic datasets.

¹email: marie@sep.Stanford.EDU,bob@sep.Stanford.EDU,biondo@sep.Stanford.EDU

THEORY

Migration in the reflection angle domain

Migration in the reflection angle domain is a subsurface-oriented technique. It is most easily understood as a Kirchhoff method. For each point in the subsurface (\mathbf{x}), a ray couple can be created for a certain reflection angle (θ) and a certain dip angle (ϕ) as shown in Figure 1. To create a CIG, simply sum over all of the dip angles:

$$a(\mathbf{x}, \theta) = \int w(\phi, \theta, \mathbf{x}) d(s, r, t) d\phi \quad (1)$$

where $a(\mathbf{x}, \theta)$ is the RAD CIGs, $w(\phi, \theta, \mathbf{x})$ is an appropriate weighting function, and $d(s, r, t)$ is the data in source, receiver and time space. From this, it is clear that each event in a RAD CIG is created by one and only one ray couple. This means there is no multipathing. Therefore, we will have fewer artifacts, especially in complex areas. In this paper, we actually use a wave-

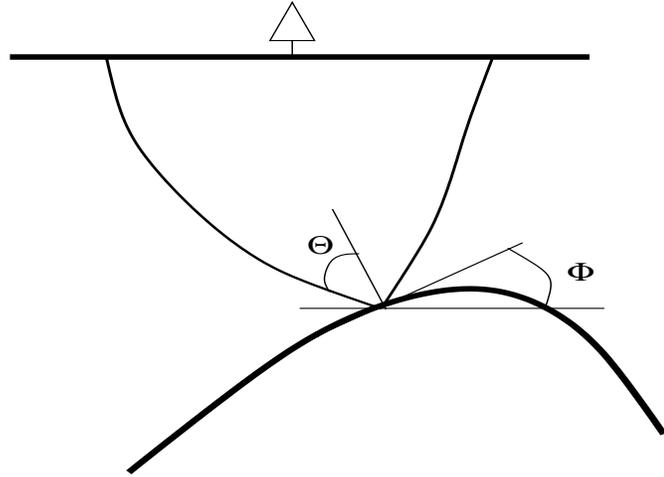


Figure 1: Geometry of the reflection angle marie1-rangle [NR]

equation method to create the RAD CIG. We begin by using the Double Square Root (DSR) equation to downward continue the data (represented as the wavefield recorded at the surface $P(\omega, \mathbf{m}, h_x; z = 0)$ where ω indicates the frequency domain, \mathbf{m} is the midpoint location, h_x is offset, and z is depth), then we slant stack (transforming the data into vertical travelttime space τ) at each depth level and image at zero time:

$$P(\omega, \mathbf{m}, h_x; z = 0) \xrightarrow{\text{DSR}} P(\omega, \mathbf{m}, h_x; z) \quad (2)$$

$$P(\omega, \mathbf{m}, h_x; z) \xrightarrow{\text{Slant stack}} P(\tau, \mathbf{m}, p_{hx}; z) \quad (3)$$

$$P(\tau, \mathbf{m}, p_{hx}; z) \xrightarrow{\text{Imaging}} P(\tau = 0, \mathbf{m}, p_{hx}; z). \quad (4)$$

The actual CIG is extracted at a fixed midpoint. The resulting CIG has an axis that is in offset ray parameter (p_{hx}) rather than reflection angle (θ). However, the offset ray parameter is related to the reflection angle by:

$$\frac{\partial t}{\partial h} = p_{hx} = \frac{2 \sin \theta \cos \phi}{V(z, \mathbf{m})}. \quad (5)$$

Sava and Fomel (2000) describe an alternative method of obtaining RAD CIGs in the Fourier domain.

The advantages of the RAD in helping to reduce illumination problems are not immediately obvious. Imagine upward continuing a wavefield from a reflector below a complex structure such as a salt dome. As the wavefield passes through the complex structure, the energy gets spread out unevenly over a large distance. In the offset and shot domains, this distribution can lead to the loss of energy that is present but doesn't conform to the regular surface-oriented geometry. The subsurface-oriented reflection angle domain does not have this difficulty. Unfortunately, there is another problem. Our acquisition geometry is always finite, therefore we can never record the entire wavefield at the surface. When we downward continue the finite recorded wavefield, the "missing" data creates a model null space. This null space cannot be filled with migration alone, so we turn to inversion.

Inversion in the reflection angle domain

Since we are using this method to image complex areas, just performing migration is not enough (Prucha et al., 1999). We want to formulate the process as an inversion problem (Nemeth et al., 1999; Duquet and Marfurt, 1999). Unfortunately, the complexity can cause the result of the inversion to diverge. Therefore, to constrain the inversion to a reasonable result, we choose to impose regularization. This can be represented by these fitting goals:

$$\mathbf{d} \approx \mathbf{Lm} \quad (6)$$

$$0 \approx \epsilon \mathbf{Am} \quad (7)$$

where \mathbf{d} is the data space, \mathbf{m} is the model space, \mathbf{L} is the wave-equation operator, \mathbf{A} is the regularization operator, and choosing the parameter ϵ is a Lagrange multiplier that allows us to determine how strong our regularization (smoothing) will be.

To speed the convergence of the inversion, we can reformulate the inversion problem as a preconditioning problem with a preconditioning operator \mathbf{S} . This operator should be as close to the inverse of the regularization operator as possible so that $\mathbf{AS} \approx \mathbf{I}$. By mapping the multi-dimensional regularization operator \mathbf{A} to helical space and applying polynomial division, we can obtain the exact inverse so that $\mathbf{S} = \mathbf{A}^{-1}$ (Claerbout, 1998). We also use the preconditioning transformation $\mathbf{m} = \mathbf{Sp}$ (Fomel et al., 1997). Our equations then become:

$$\mathbf{d} \approx \mathbf{LSp} \quad (8)$$

$$0 \approx \epsilon \mathbf{p}. \quad (9)$$

The regularization we choose for RAD inversion is horizontal smoothing along the reflection angle (or p_{hx}) axis. This is based on both theory and practice. In theory, the regularization operator should be related to the inverse model covariance matrix (Tarantola, 1986). In practice, because the covariance matrix of a CIG created with the correct velocity model is horizontal in nature (Figure 2) and the reflectivity along reflection angles should vary smoothly (Richter, 1941), we simply smooth horizontally along the reflection angle axis. An operator that fulfills

these requirements is a steering filter (Clapp et al., 1997). This steering filter and its impulse response is in Figure 3. The coefficients in the second column of the steering filter are variable and control the vertical width of the impulse response.

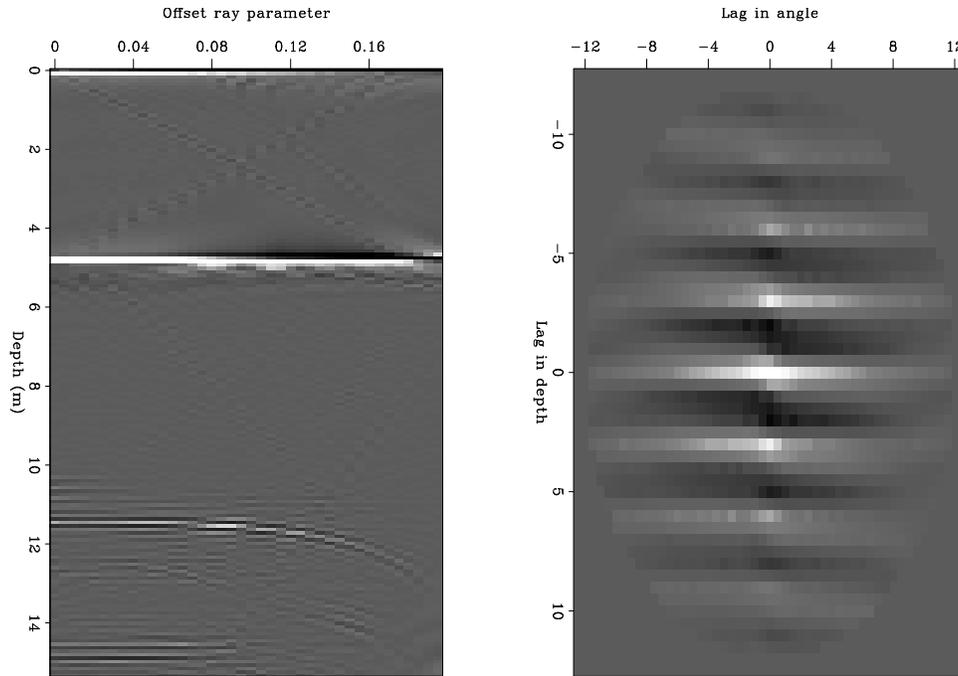


Figure 2: Left: a typical RAD CIG with some minor artifacts. Right: representation of the covariance matrix calculated for this CIG. `marie1-covar` [ER]

We also must choose an ϵ . A large ϵ means that the regularization will be strong (we will be smoothing more) and a small ϵ means the regularization is not strong (we honor the model more). For now, we are trying to choose an ϵ that is somewhere in the middle, so that our result is slightly smoothed. We want to try to fill the null space in a reasonable way.

RESULTS

Elf North Sea dataset

The first dataset we applied this RAD preconditioned inversion to was an Elf North Sea dataset. It is a 2-D synthetic created based on a real 3-D survey. Its velocity model can be seen in Figure 4. This dataset is known to have illumination problems under the edge of the salt (Prucha et al., 1998). We first created CIGs by just RAD migration, then we created CIGs by preconditioned inversion. The inversion results are from just one iteration.

Figure 5 shows a comparison of the results from a CMP location that is in a simple area of the subsurface. Note that the inversion result is cleaner, smoother, and more continuous than the migrated result. A good example is the event at 4400 meters, which has been cleaned up

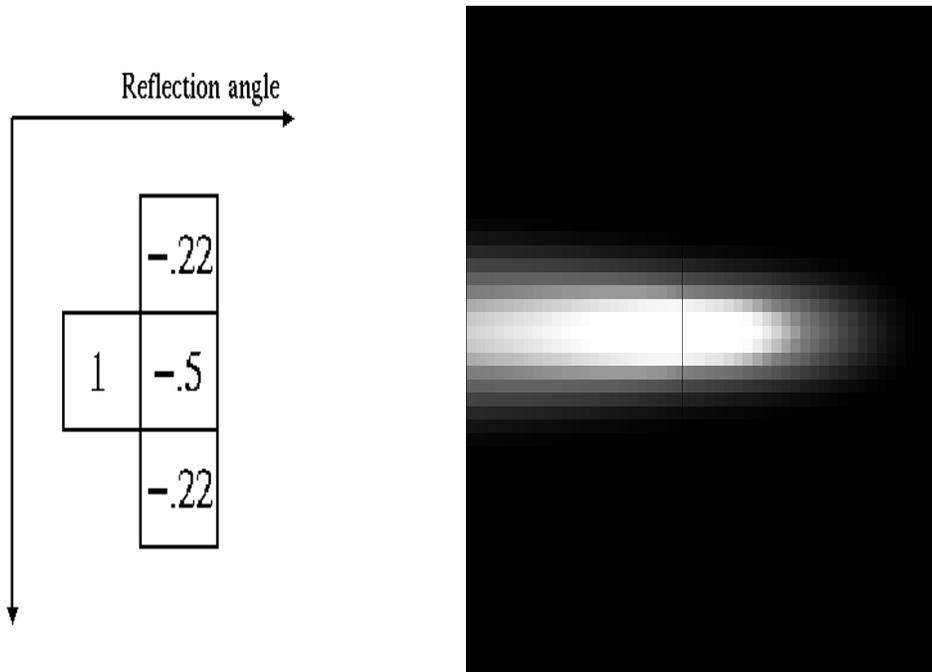


Figure 3: Left: steering filter used for regularization. Right: impulse response of steering filter. `marie1-filtres` [NR]

and now covers almost twice as many reflection angles as it did in the migrated result while still being reasonable.

Figure 6 displays the CIGs for a CMP at the edge of the salt. Once again the inversion result is better than the migrated result, and even the event for the poorly illuminated reflector at depth 4400 meters is more continuous.

BP multiple dataset

We also applied this method to a synthetic multiple dataset provided to us by BP. This is a complicated dataset designed to have extremely nasty multiples. Its velocity model is in Figure 7. Once again, we can expect illumination problems under the edges of the salt body (Muerdter et al., 1996). Even after only one iteration, the RAD preconditioned inversion has produced CIGs that are cleaner and more continuous than RAD migration.

Figure 8 displays a CIG taken from a point away from the salt body. The migrated result has good, continuous reflectors, but there are artifacts between 6 kilofeet and 10 kilofeet and there is a multiple at depth 11.5 kilofeet. The inversion result cleans these artifacts up and attenuates the multiple.

Figure 9 shows a CIG taken at the right edge of the salt body. There are many artifacts and multiples in both the migrated and the inversion results, but the strength of the artifacts and the multiples is reduced in the preconditioned result. Preconditioning has also made the event

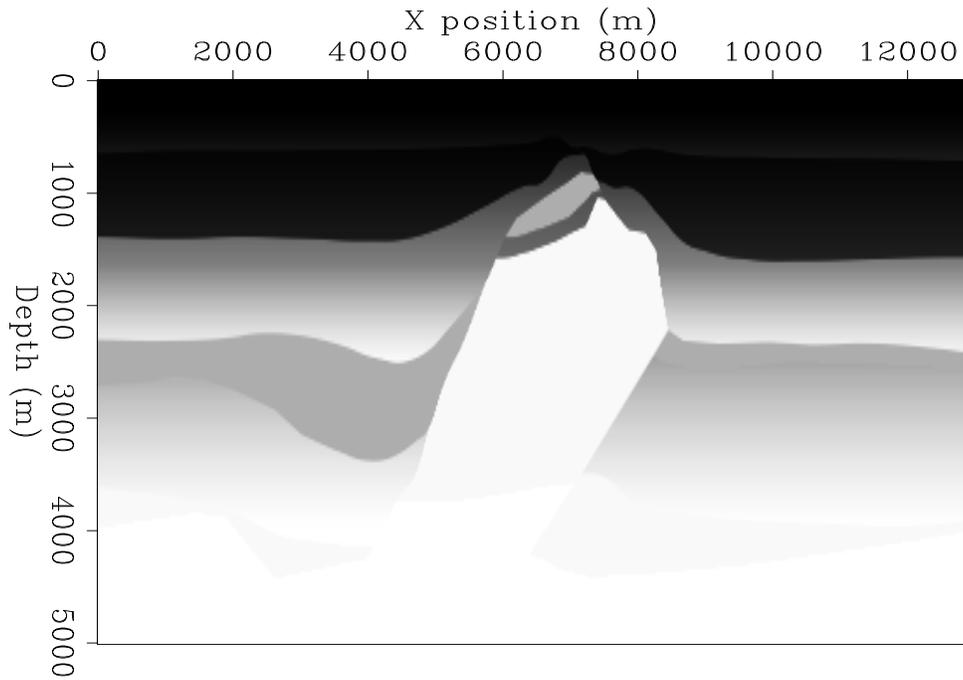


Figure 4: Velocity model of the Elf North Sea dataset. `marie1-elfvel` [ER]

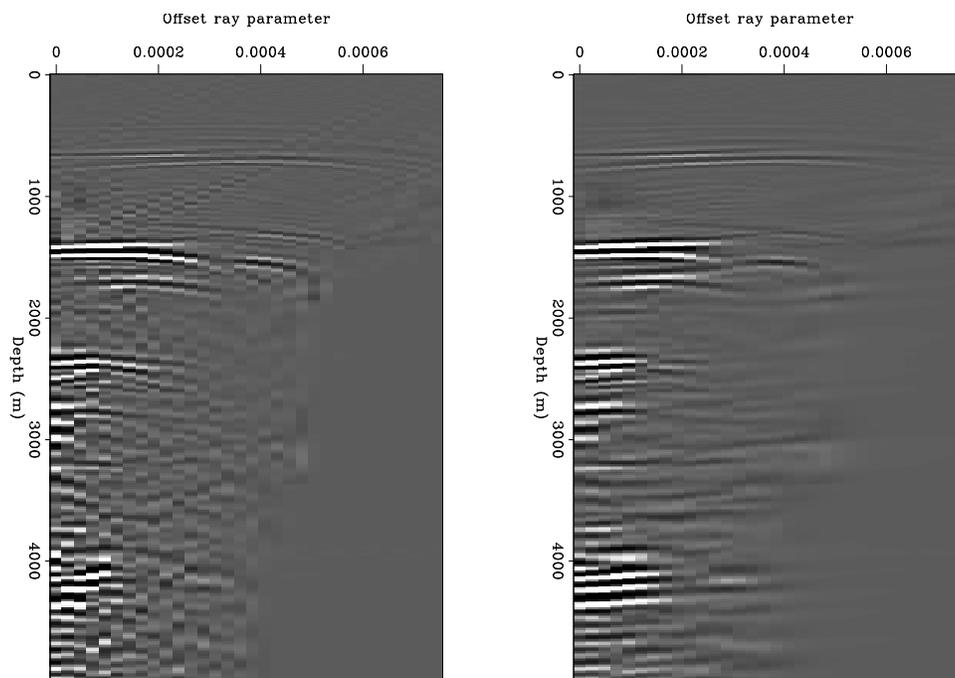


Figure 5: Left: RAD CIG from migration only at CMP 200. Right: RAD CIG from inversion at CMP 200. `marie1-elf.comp200` [CR]

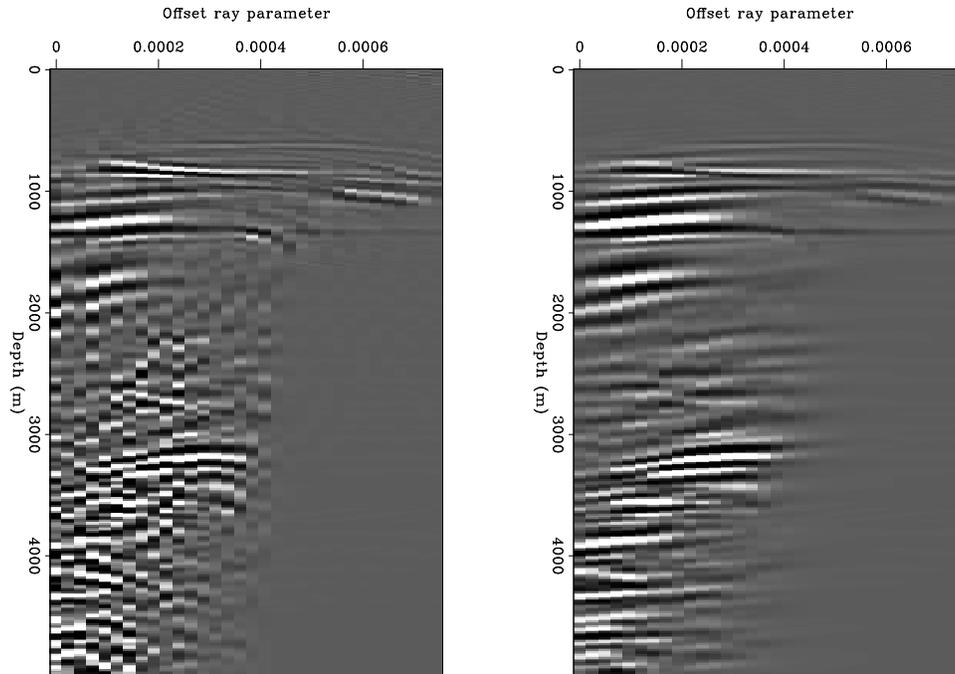


Figure 6: Left: RAD CIG from migration only at CMP 7300. Right: RAD CIG from inversion at CMP 7300. marie1-elf.comp7300 [CR]

from the reflector below the salt at depth 13 kilofeet more continuous.

CONCLUSIONS

Regularization in the reflection angle domain helps clean up the artifacts in the common image gathers and improves the continuity of the events in the CIGs. The use of steering filters as our regularization operator is a simple and effective choice. The additional cost of inversion is worthwhile in complex areas.

ACKNOWLEDGMENTS

We would like to thank Elf and IFP for the Elf North Sea synthetic dataset and BP-Amoco for the BP multiple dataset.

REFERENCES

- Chemingui, N., 1999, Imaging irregularly sampled 3D prestacked data: Ph.D. thesis, Stanford University.
- Claerbout, J. F., 1991, Design of inverse Kirchhoff-style filters by LSCG: SEP-72, 33-38.

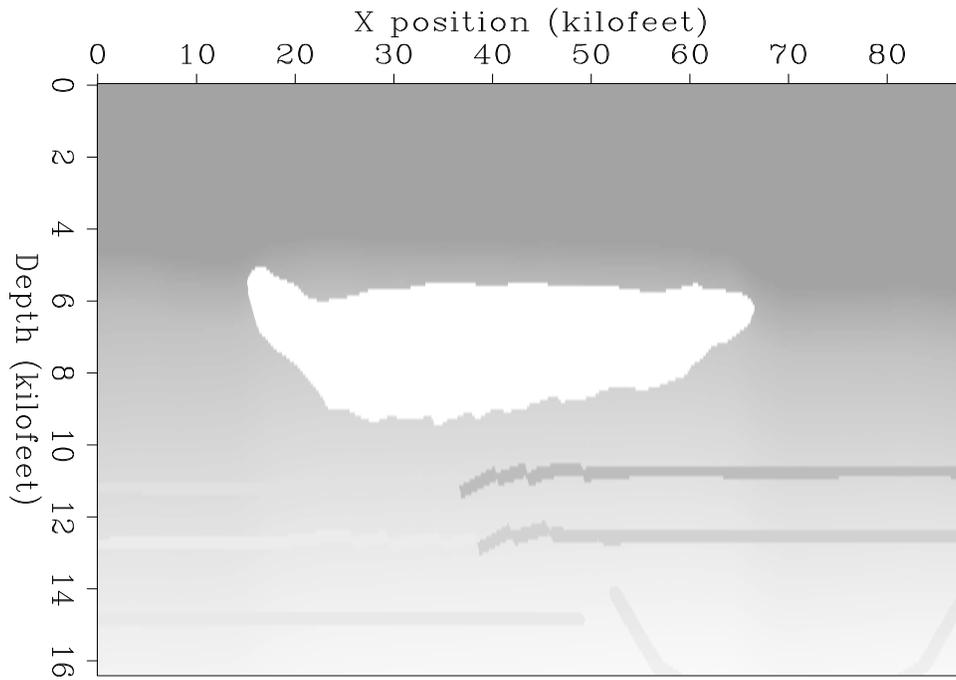


Figure 7: P-wave velocity model of the BP multiple dataset. `marie1-bpvel` [ER]

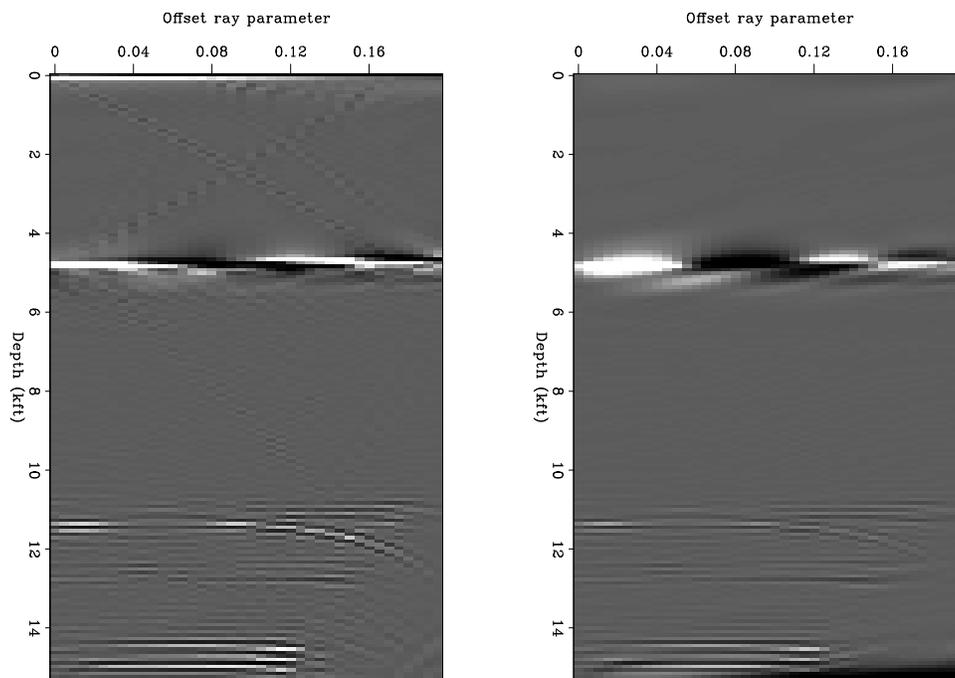


Figure 8: Left: RAD CIG from migration only at CMP 10. Right: RAD CIG from inversion at CMP 10. `marie1-BP.comp10` [CR]

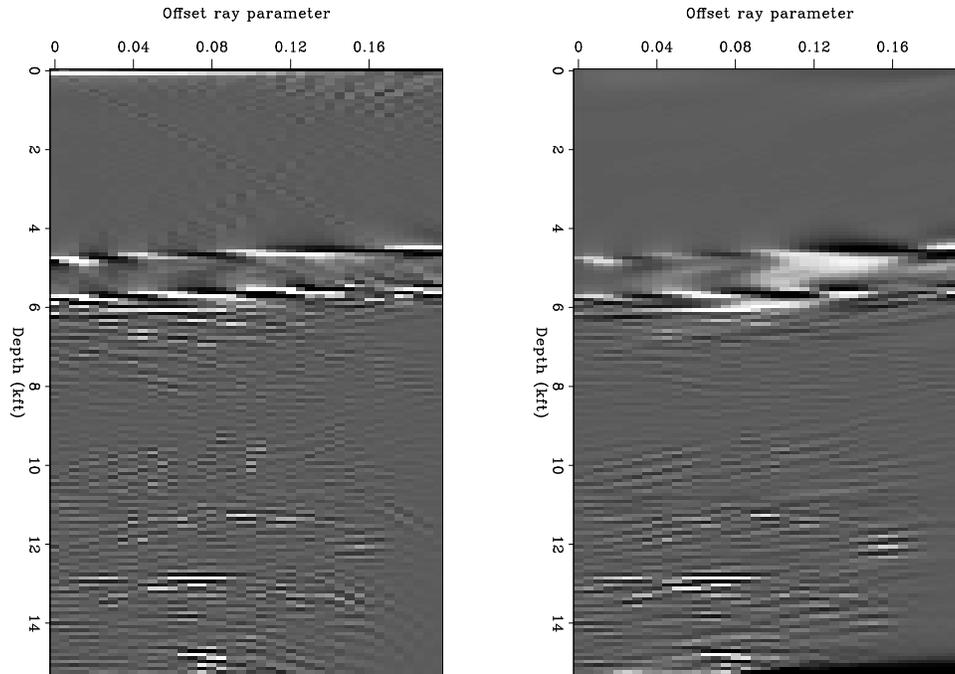


Figure 9: Left: RAD CIG from migration only at CMP 66.33. Right: RAD CIG from inversion at CMP 66.33. `marie1-BP.comp66.33` [CR]

Claerbout, J., 1998, Multidimensional recursive filters via a helix: *Geophysics*, **63**, no. 05, 1532–1541.

Clapp, R. G., Fomel, S., and Claerbout, J., 1997, Solution steering with space-variant filters: *SEP-95*, 27–42.

Duquet, B., and Marfurt, K. J., 1999, Filtering coherent noise during prestack depth migration: *Geophysics*, **64**, no. 4, 1054–1066.

Fomel, S., Clapp, R., and Claerbout, J., 1997, Missing data interpolation by recursive filter preconditioning: *SEP-95*, 15–25.

Fomel, S., 1997, On model-space and data-space regularization: A tutorial: *SEP-94*, 141–164.

Harlan, W. S., 1986, Signal-noise separation and seismic inversion: *SEP-47*.

Muerdter, D. R., Lindsay, R. O., and Ratcliff, D. W., 1996, Imaging under the edges of salt sheets: a raytracing study: 66th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 578–580.

Nemeth, T., Wu, C., and Schuster, G. T., 1999, Least-squares migration of incomplete reflection data: *Geophysics*, **64**, no. 1, 208–221.

Prucha, M. L., Clapp, R. G., and Biondi, B. L., 1998, Imaging under the edges of salt bodies: Analysis of an Elf North Sea dataset: *SEP-97*, 35–44.

- Prucha, M., Biondi, B., and Symes, W., 1999, Angle-domain common image gathers by wave-equation migration: 69th Ann. Internat. Meeting, Soc. Expl. Geophysics, Expanded Abstracts, 824–827.
- Richter, C. F., 1941, *Elementary Seismology*: Freeman, San Francisco.
- Sava, P., and Fomel, S., 2000, Angle-gathers by Fourier Transform: SEP-103, 119–130.
- Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection data: *Geophysics*, **51**, no. 10, 1893–1903.
- ten Kroode, A. P. E., Smit, D.-J., and Verdel, A. R., 1999, Linearized inverse scattering in the presence of caustics: *Wave Motion*.
- Xu, S., Chauris, H., Lambare, G., and Noble, M., 1998, Common angle image gather - a strategy for imaging complex media: Summer Research Workshop: Depth Imaging of Reservoir Attributes, EAGE-SEG, Expanded Abstracts, X012.

Angle-gathers by Fourier Transform

Paul Sava and Sergey Fomel¹

ABSTRACT

In this paper, we present a method for computing angle-domain common-image gathers from wave-equation depth-migrated seismic images. We show that the method amounts to a radial-trace transform in the Fourier domain and that it is equivalent to a slant stack in the space domain. We obtain the angle-gathers using a stretch technique that enables us to impose smoothness through regularization. Several examples show that our method is accurate, fast, robust, easy to implement and that it can be used for real 3-D prestack data in applications related to velocity analysis and amplitude-versus angle (AVA) analysis.

INTRODUCTION

Traditionally, migration velocity analysis and AVO employ offset-domain common-image gathers, since most of the relevant information is not described by the zero-offset images. However, it is difficult to produce these gathers with wave-equation migration because the offset dimension of the downward continued data shrinks with depth. A solution to this problem is to use angle-gathers instead of offset-gathers. Angle-gathers also provide more straightforward information for amplitude analysis, that is, amplitude variation with angle (AVA) instead of the more common amplitude variation with offset (AVO).

Angle-domain common-image gathers are representations of the seismic images sorted by the aperture angle at the reflection point (Prucha et al., 1999a; Fomel and Prucha, 1999). In the space-domain, the reflection angles can be evaluated by the simple differential equation:

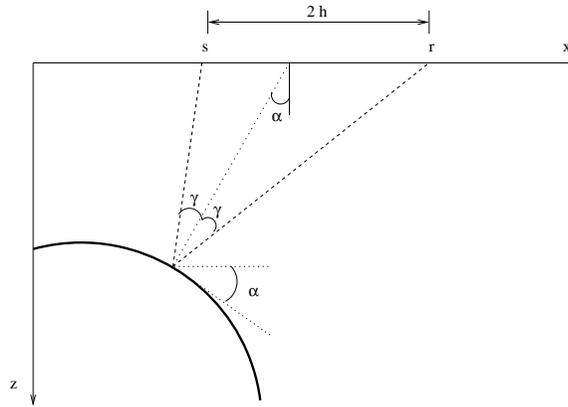
$$\tan \gamma = -\frac{\partial z}{\partial h}$$

where γ represents half of the aperture angle at the reflector, and z and \vec{h} are, respectively, the depth and half-offset for the particular reflection (Figure 1).

Prucha et al. (1999a) evaluate wave-equation angle-domain common-image gathers (AD-CIG) by slant-stacking between the downward continuation and the imaging steps. They also produce their output as a function of offset ray-parameter (p) instead of the reflection angle (γ). However, angle-domain gathers do not need to be computed by slant stacks directly, but could be more easily evaluated in the Fourier-domain, with output in the true reflection angle,

¹email: paul@sep.stanford.edu, sergey@sep.stanford.edu

Figure 1: The geometrical relationship relating the reflection angle γ at the reflector, to the half-offset \bar{h} .
paul2-rays [NR]



using the simple equation:

$$\tan \gamma = -\frac{|\vec{k}_h|}{k_z}. \quad (1)$$

which, in essence, is simply a stretch of the offset axis transformed in the Fourier-domain.

It is interesting to emphasize that this Fourier-domain method of computation of angle-gathers is a radial-trace transform. Ottolini (1982) shows that slant stacks are, in fact, generally equivalent to radial-trace transforms, with significant advantages and disadvantages for each of these two alternatives.

In this paper, we review the radial stretch theory and show that computing the ADCIGs using equation (1) is indeed equivalent to slant stacks but that it gives us the freedom to postpone the conversion to angle until after prestack wave-equation migration. This Fourier domain method makes it possible to convert seismic images easily from the offset-domain to the angle-domain without re-migrating the data.

The novelty of our method is that, in the Fourier-domain, the radial-trace transformation can be easily regularized, thus leading to smooth and continuous representations of the reflectors along the angle axis. A similar method of regularized stretch has been used in other applications: velocity continuation (Fomel, 1998), Stolt migration (Vaillant and Fomel, 1999), and Stolt residual migration (Sava, 1999b,a). It could also find many other applications, for example in coherent-noise removal using radial-trace transforms (Brown and Claerbout, 2000).

Our method of angle-gather computation is different from the slant-stack approach not only in the domain in which it operates, but also in the low cost involved, the robustness, and simplicity of the implementation. If we compute angle-gathers in conjunction with Stolt residual migration, this method becomes even more attractive since the seismic images are already transformed to the Fourier-domain, which makes the true cost of the transformation insignificant.

EQUIVALENCE TO SLANT STACKS

The Fourier-domain stretch represented by equation (1) is equivalent to a slant stack in the $z - \vec{h}$ domain. Indeed, we can convert an image gather in the offset-domain (\mathbf{H}) to one in the angle-domain (\mathbf{A}), using a slant-stack equation of the form

$$\mathbf{A}(z, \vec{\mu}) = \int \mathbf{H}(z + \vec{\mu} \cdot \vec{h}, \vec{h}) d\vec{h}, \quad (2)$$

where $\vec{\mu}$ is a vector describing the direction of the stack.

Fourier transforming equation (2) over the depth axis, we obtain

$$\underline{\mathbf{A}}(k_z, \vec{\mu}) = \int \left[\int \mathbf{H}(z + \vec{\mu} \cdot \vec{h}, \vec{h}) d\vec{h} \right] e^{ik_z z} dz$$

where the underline stands for a 1-D Fourier transform. We can continue by writing the equation

$$\underline{\mathbf{A}}(k_z, \vec{\mu}) = \int \int \mathbf{H}(z + \vec{\mu} \cdot \vec{h}, \vec{h}) e^{ik_z(z + \vec{\mu} \cdot \vec{h}) - ik_z \vec{\mu} \cdot \vec{h}} d\vec{h} dz,$$

where we can re-arrange the terms as

$$\underline{\mathbf{A}}(k_z, \vec{\mu}) = \int \left[\int \mathbf{H}(z + \vec{\mu} \cdot \vec{h}, \vec{h}) e^{ik_z(z + \vec{\mu} \cdot \vec{h})} dz \right] e^{-ik_z \vec{\mu} \cdot \vec{h}} d\vec{h},$$

which highlights the relation between the 1-D Fourier-transformed angle-domain and offset-domain representation of the seismic images:

$$\underline{\mathbf{A}}(k_z, \vec{\mu}) = \int \underline{\underline{\mathbf{H}}}(k_z, \vec{h}) e^{-ik_z \vec{\mu} \cdot \vec{h}} d\vec{h}.$$

We recognize on the right-hand side of the previous equation additional Fourier transforms over the offset axes, and therefore we can write

$$\underline{\mathbf{A}}(k_z, \vec{\mu}) = \underline{\underline{\underline{\mathbf{H}}}}(k_z, -\vec{\mu}k_z),$$

where the triple underline stands for the 3-D Fourier transform of the offset-domain common-image gather. Finally, defining $-\vec{\mu}k_z = \vec{k}_h$, we can conclude that the 1-D Fourier transforms of angle-domain gathers are equivalent to the 3-D Fourier transforms of the offset-domain gathers,

$$\underline{\mathbf{A}}(k_z, \vec{\mu}) = \underline{\underline{\underline{\mathbf{H}}}}(k_z, \vec{k}_h), \quad (3)$$

subject to the stretch of the offset axis according to the simple law

$$\vec{\mu} = -\frac{\vec{k}_h}{k_z}. \quad (4)$$

We can recognize in equation (4) the fundamental relation between the reflection angle and the Fourier-domain quantities that are evaluated in wave-equation migration. This equation also shows that the angles evaluated by (1) are indeed equivalent to slant stacks on offset-domain common-image gathers. Therefore, we could either compute angles for each of the two offset axes with the equations

$$\begin{aligned}\gamma_x &= -\tan^{-1}\left(\frac{k_{hx}}{k_z}\right) \\ \gamma_y &= -\tan^{-1}\left(\frac{k_{hy}}{k_z}\right),\end{aligned}$$

or compute one angle corresponding to the entire offset vector:

$$\gamma = -\tan^{-1}\left(\frac{|\vec{k}_h|}{k_z}\right).$$

REGULARIZATION OF THE ANGLE DOMAIN

In essence, the angle-gather method, introduced in this paper, amounts to a stretch of the offset angle according to equation (4). The stretch takes every point on the offset wavenumber axis and repositions it on the angle axis, most likely not on its regular grid. We therefore need to interpolate the unevenly sampled axis to the regular one. In other words, we need to solve a simple linear interpolation problem

$$\mathbf{Lm} \approx \mathbf{d}$$

where the model (\mathbf{m}) is represented by the evenly-spaced values on the angle axis, the data (\mathbf{d}) is represented by the unevenly-spaced values on the angle axis, and (\mathbf{L}) represents a 1-D linear interpolation operator. Since parts of the model space will not be covered because of the uneven distribution of the data, we need to regularize the interpolation process and solve a system such as

$$\mathbf{Lm} \approx \mathbf{d} \tag{5}$$

$$\epsilon \mathbf{Am} \approx 0 \tag{6}$$

where (\mathbf{A}) represents a 1-D roughener operator. Consequently, the least-squares solution to the system (5) is

$$\mathbf{m} = (\mathbf{L}^T \mathbf{L} + \epsilon^2 \mathbf{A}^T \mathbf{A})^{-1} \mathbf{L}^T \mathbf{d}. \tag{7}$$

In the special case of the angle-domain stretch, the inverted term on the right side of equation (7) is a tridiagonal matrix. Given the sparseness of the stretched data, the least-squares tridiagonal matrix corresponding to the operator \mathbf{L} has zeros present along the diagonals, which results in instability during inversion. However, the regularization term fills the gaps; therefore, the inversion of the matrix in equation (7) is well-behaved.

Since the matrix $\mathbf{L}^T\mathbf{L} + \epsilon^2\mathbf{A}^T\mathbf{A}$ is tridiagonal, we can invert it using a fast tridiagonal solver (Golub and Van Loan, 1989); Consequently, we obtain smoothly interpolated values for the ADCIGs. A similar approach could also be used for other problems, for example in Stolt migration (Vaillant and Fomel, 1999), residual migration (Sava, 1999b,a), or in velocity continuation (Fomel, 1998).

The main benefit of solving the least-squares problem this way is that we can obtain a very inexpensive regularized solution, with important benefits not only in data visualization, but also in other problems such as wave-equation migration velocity analysis (Biondi and Sava, 1999; Sava and Biondi, 2000) and imaging (Prucha et al., 1999b).

EXAMPLES

We exemplify the proposed method on two synthetic models and two real datasets.

The first example is a 2-D synthetic model with dipping reflectors at various angles. We generated the synthetic data using wave-equation modeling (Biondi, 1999). Next, we imaged the data, first using the correct and then using an incorrect velocity model, a slower velocity in this example. In the case of correct velocity, the ADCIGs are flat, but they are not flat in the case of the incorrect velocity model (Figure 2). Because the simulated acquisition is represented by wide offsets and the model is reasonably shallow, there is no significant decrease in the angular coverage at the deeper reflectors. However, the steep reflectors are characterized by smaller angular coverage due to the limited acquisition geometry.

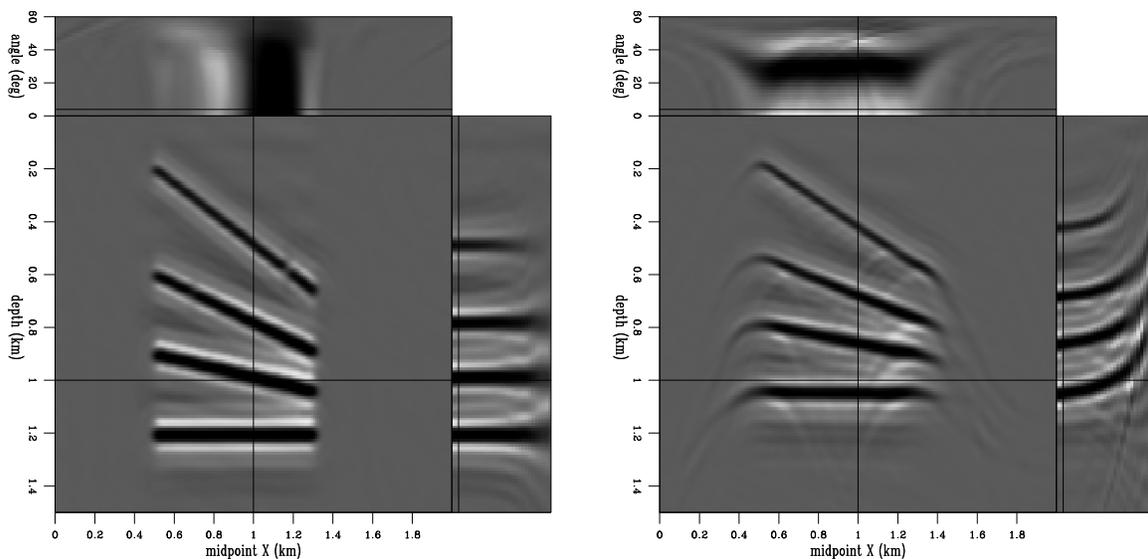


Figure 2: The first synthetic model. The reflectivity is composed of interfaces dipping at different angles. All the events are flat along the reflection angle axis, in the case where we use the correct velocity (left). However, when we use a wrong velocity model, the image loses focus and the ADCIGs bend (right). [paul2-dipsynt.cig01](#) [CR]

In the second example, we consider a more complex synthetic model, centered on a salt body. The model was generated at Elf-IFP-CGG and inspired by real data recorded in the North Sea (Prucha et al., 1998). Again, we image both with a correct velocity model and an incorrect one (Figure 3). The conclusions are similar to those in the first example: correct velocity flattens the events in the angle-domain common-image gathers, while incorrect velocity does not. Again, there is no significant decrease in the angular coverage at the deeper reflector.

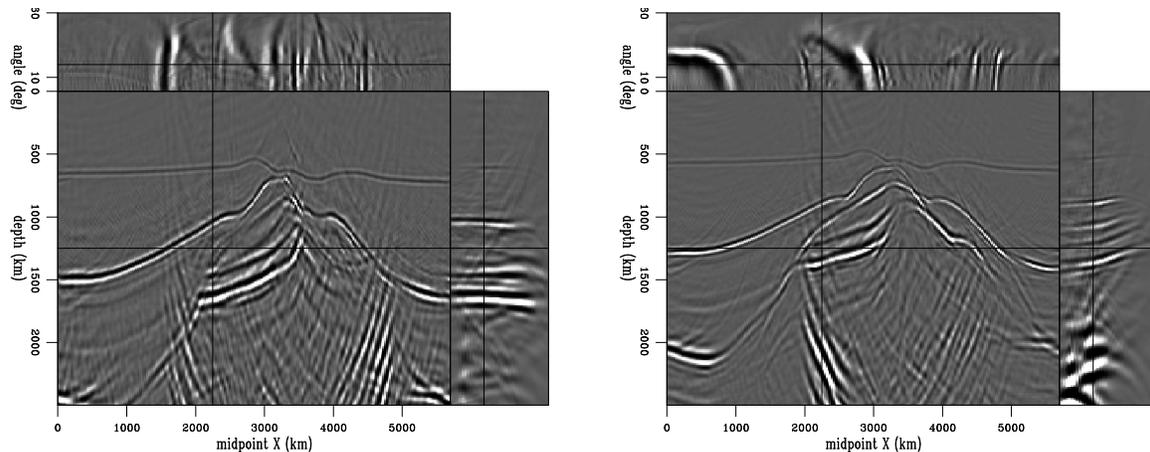


Figure 3: The second synthetic model, depicting a salt body. The ADCIGs show flat events when the velocity model is correct (left), but bend when the velocity is incorrect.

`paul2-saltsynt.cig01` [CR]

The third example is a real dataset acquired in a salt-dome region of the North Sea (Prucha et al., 1998; Vaillant and Sava, 1999). Figure 4 depicts a slice of the image taken at a small reflection angle, and the small panels at the top and the right represent ADCIGs. The image generated for this dataset goes much deeper compared to the preceding synthetics, and we see that the angular coverage decreases as the depth increases. Some of the events appear flat, while others are bending, indicating velocity inaccuracies.

It is apparent from the image that some of the events lose their sharpness and the relative contrast decreases. This is understandable, since the energy of every event is spread along the angle axis. The true migrated image could be recovered by summation along the angle axis (Figure 5).

This example enables us to analyze the efficiency of our regularization method. Figure 6 shows angle-domain common-image gathers for a particular midpoint. The left panel represents the result obtained without regularization ($\epsilon = 0.0$). In contrast, the panel on the right shows the result we obtain with regularization ($\epsilon = 1.0$). The regularized image is much cleaner, without visible distortions of the shape and amplitudes of the events. Since the noise level is much lower, we can easily identify faint events that would otherwise be impossible to discern, for example at depths greater than 4000 m.

Finally, a fourth example addresses the amplitude variation with angle analysis (AVA) is-

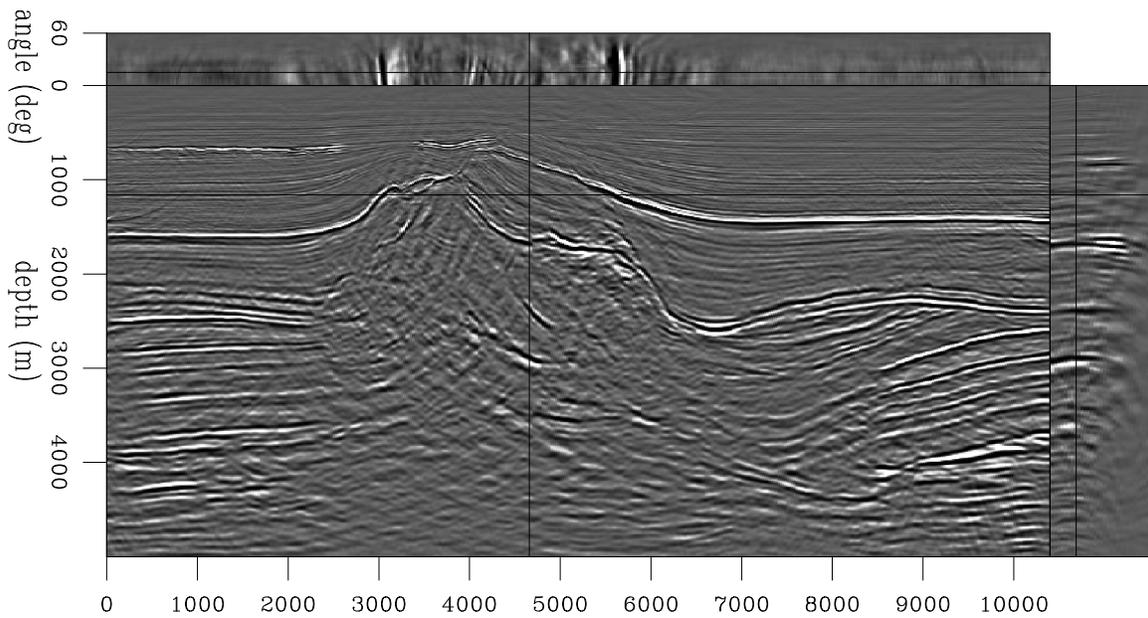


Figure 4: Migrated image for the North Sea dataset presented as ADCIGs. The ADCIGs are mostly flat, but they reveal the areas where the velocity map needs improvement. The image as a whole loses part of its sharpness, because the energy of each event is spread along the angle axis. `paul2-saltreal.cig` [CR]

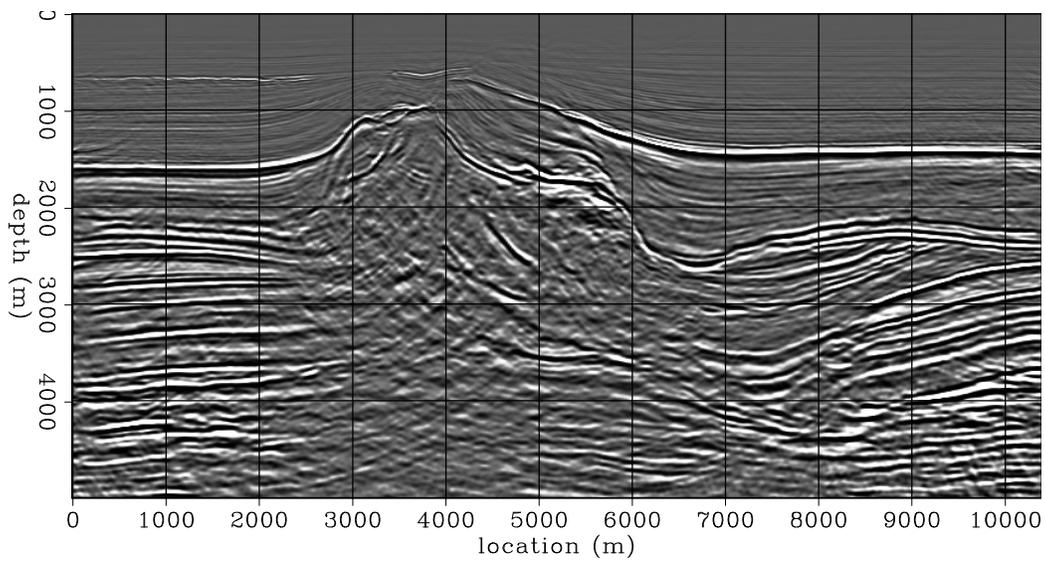
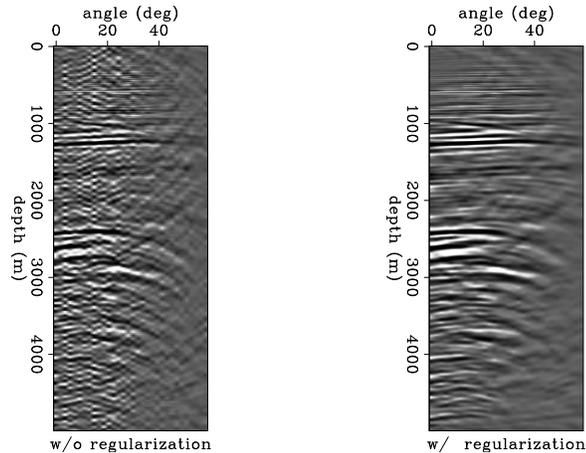


Figure 5: Migrated image for the North Sea dataset, stacked over the angle axis in the ADCIGs. `paul2-saltreal.stk` [CR]

Figure 6: A comparison of ADCIGs with regularization (right panel) and without regularization (left panel).

`paul2-saltreal.eps` [CR]



sue. Our method produces the output as a function of the reflection angle, which makes it an ideal tool for AVA analysis. The crucial question is how much the amplitudes are affected by the method, especially given that we impose a certain amount of smoothness through regularization. Previous research (Prucha et al., 1999a) indicates that ADCIGs obtained by wave-equation imaging have the potential to preserve AVA effects.

For the AVA example, we use a dataset from a gas-hydrates study, for which the AVO effects are significant (Ecker, 1998). Figure 7 represents the wave-equation migrated image, displayed as ADCIGs. Figure 8 shows 4 wiggle-plots of ADCIGs located around 45 km from the left origin of the survey. This area is particularly interesting since it captures both the bottom-simulation reflector (BSR) and the flat reflector from the hydrate-gas transition. The results confirm the amplitude trend variation previously highlighted for this dataset; specifically, we can observe significant amplitude increase as a function of angle, especially for the BSR. This result was obtained for a moderate choice of the regularization parameter ($\epsilon = 1.0$). Caution should be exercised in this matter, since a higher value could attenuate the amplitude of the variations. A more thorough and quantitative analysis awaits future research.

It is worth mentioning that, in certain applications, the amplitude variations with angle are undesirable. A good example is velocity-analysis where AVA effects can be misleading and indicate incorrect updates of the velocity model. However, we can perhaps control this by boosting the ϵ parameter, with the effect of attenuating the variations along angles.

DISCUSSION

Several key points highlight the strengths of our method of computing angle-domain common-image gathers:

- Our method produces the output in the reflection angle at the reflector, and not in the offset ray-parameter. This makes the results more open to interpretation, and potentially allows for consistent quantitative AVA analysis.

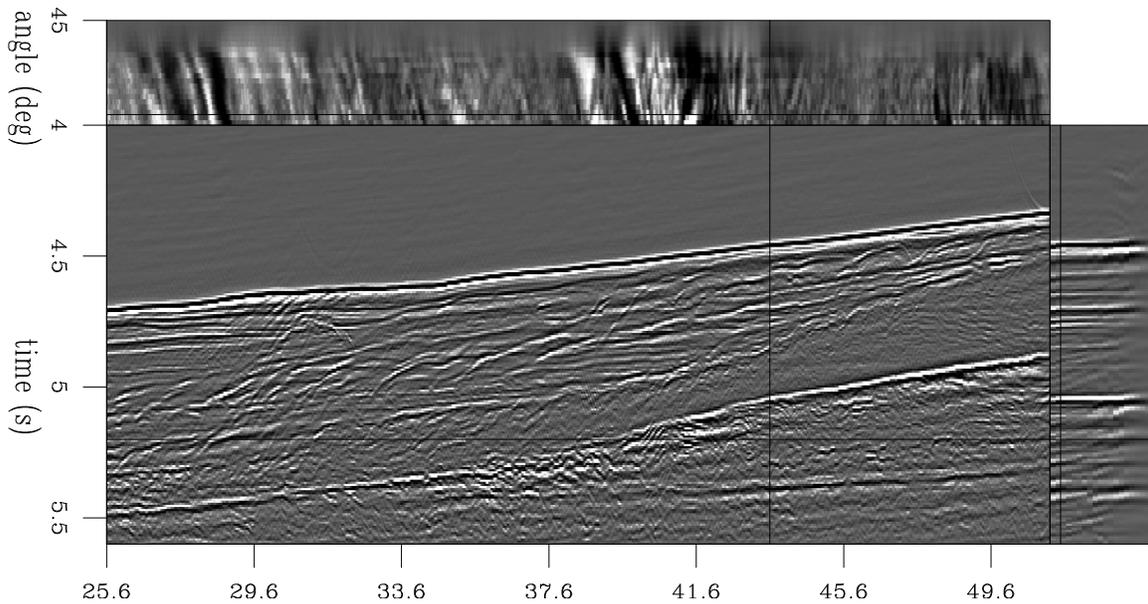


Figure 7: Migrated image for the gas-hydrates data presented as ADCIGs. `paul2-hydrreal.cig`
[CR]

- Our method generates angle gathers after and not during migration, thus enabling us to shuttle between the angle and offset domains without re-migrating the data. In addition, the method is much easier to code, and therefore potentially more robust.
- Our method enables inexpensive regularization of the angle-domain, leading to gathers with events that vary smoothly along the angle axis. The increased S/N ratio helps reveal weak events that would otherwise be impossible to find.
- If used in conjunction with prestack Stolt residual migration, the cost of our method becomes trivial, since the images are already transformed to the Fourier-domain. Also, returning to the space-domain after our transformation is less costly, since we reduce the need for a 3-D Fourier transform at every midpoint to that of a 1-D Fourier transform.

CONCLUSIONS

We have presented a method for computing angle-domain common-image gathers from wave-equation depth-migrated images. We have shown that the method is, in essence, a radial trace transform in the Fourier domain, and therefore equivalent to a slant stack in the space domain. We used a stretch technique that enabled us to include model regularization, which leads to smooth ADCIGs. We have also shown that the method is accurate, fast, robust, easy to implement and that it can be used for real 3-D prestack data in applications related to velocity and AVA analysis.

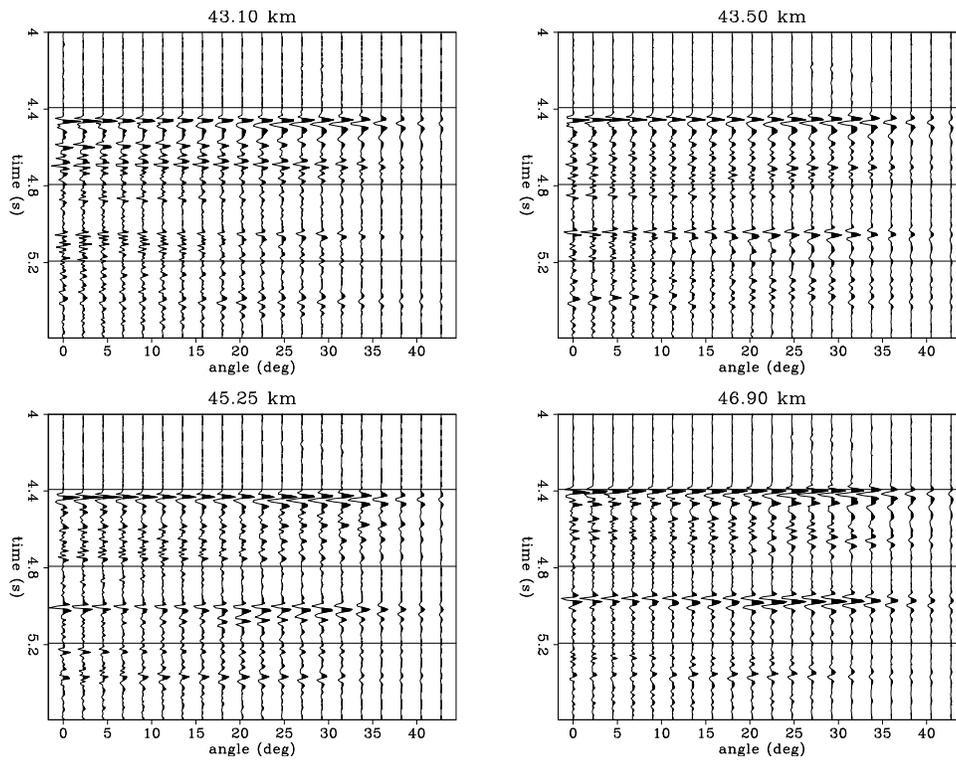


Figure 8: AVA analysis panels for the hydrates data. The BSR (located around 4.4 s) clearly shows consistent increase of the amplitude with angle and confirms previously reported results. Each panel is labeled with the horizontal coordinate of the image-gather. `paul2-hydrreal.ava` [CR]

ACKNOWLEDGMENT

Elf Aquitaine provided both the synthetic salt-dome model and the North Sea dataset.

REFERENCES

- Biondi, B., and Sava, P., 1999, Wave-equation migration velocity analysis: 69th Ann. Internat. Meeting, Soc. Expl. Geophys., Expanded Abstracts, 1723–1726.
- Biondi, B. L., 1999, 3-D Seismic Imaging: Stanford Exploration Project.
- Brown, M., and Claerbout, J., 2000, Ground roll and the Radial Trace Transform - revisited: SEP-103, 219–237.
- Ecker, C., 1998, Seismic characterization of methane hydrate structures: Ph.D. thesis, Stanford University.
- Fomel, S., and Prucha, M., 1999, Angle-gather time migration: SEP-100, 141–150.
- Fomel, S., 1998, Velocity continuation by spectral methods: SEP-97, 157–172.
- Golub, G. H., and Van Loan, C. F., 1989, Matrix computations: The Johns Hopkins Press Ltd., London.
- Ottolini, R., 1982, Migration of reflection seismic data in angle-midpoint coordinates: Ph.D. thesis, Stanford University.
- Prucha, M. L., Clapp, R. G., and Biondi, B. L., 1998, Imaging under the edges of salt bodies: Analysis of an Elf North Sea dataset: SEP-97, 35–44.
- Prucha, M., Biondi, B., and Symes, W., 1999a, Angle-domain common-image gathers by wave-equation migration: 69th Ann. Internat. Meeting, Soc. Expl. Geophys., Expanded Abstracts, 824–827.
- Prucha, M. L., Biondi, B. L., and Symes, W. W., 1999b, Angle-domain common image gathers by wave-equation migration: SEP-100, 101–112.
- Sava, P., and Biondi, B., 2000, Wave-equation migration velocity analysis: Episode II: SEP-103, 19–47.
- Sava, P., 1999a, Short note—on Stolt common-azimuth residual migration: SEP-102, 61–66.
- Sava, P., 1999b, Short note—on Stolt prestack residual migration: SEP-100, 151–158.
- Vaillant, L., and Fomel, S., 1999, On stolt stretch time migration: SEP-102, 67–76.
- Vaillant, L., and Sava, P., 1999, Common-azimuth migration of a North Sea dataset: SEP-102, 1–14.

Spectral factorization of 2-D reflection seismic data

*Jon Claerbout*¹

ABSTRACT

I propose spectral factorization of 2-D seismic data. Boulders strewn on the water bottom of an otherwise horizontally layered earth imply that the multidimensional minimum-phase wavelet of a zero-offset section is a common midpoint gather.

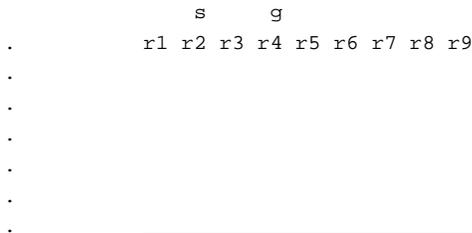
INTRODUCTION

Recently, math professor George Papanicolaou delivered a seminar to our Geophysics Department in which he presented to us an amazing proposition: Random scatterers can give us superresolution because they can enlarge the effective aperture.

I will propose it this way: Scatterers give rise to seismic coda. Maybe we can use it effectively even though we may not know the location of the scatterers.

In this paper I sketch how this could happen in the realistic case of near-surface scattering and I indicate how it could be tested and demonstrated.

Except for one essential feature, the earth model that we examine in this initial exploratory phase is a two-dimensional horizontally layered earth. The essential departure is that the top roughness that acts as point scatterers. We might think of it as fine scale surface topography. Alternately, we might think of it as a thin water layer with boulders strewn around, all acting as point scatterers of random amplitude, polarity and location. The one-dimensional earth model itself has arbitrary velocity $v(z)$, multiple reflections, shear waves, anisotropy, etc. Visualize this geometry:



We are interested in ray paths like the one from the shot s to the reflector, to the rock $r9$, to the reflector, to the geophone g . Both paths to and from the rock $r9$ include all arrivals,

¹email: claerbout@stanford.edu

both direct and reflected. We convolve the 1-D simple earth's s to r_9 response with its r_9 to g response. Whatever the response is to one rock, the response to all the rocks will be the one-rock response convolved horizontally by random numbers.

There will be several conjectures. The simplest is that from a zero-offset section, spectral factorization (Kolmogoroff, 1939) via the helix (Claerbout, 1998) manufactures a common midpoint gather from which we can do velocity analysis.

SUBTRACT ANY TWO SHOT GATHERS

We will soon be doing some data-guided theory. Carefully examine Figure 1. I have only this one shot gather (a photo image) but I'd like you to imagine two of them. Imagine the earth really is a one-dimensional layered medium with surface scatterers. Then the two shot gathers would look identical EXCEPT for the diffraction at 250m offset. This diffraction would be differently positioned on each shot gather depending on the distance of the shot from the surface scatterer. Let us subtract these two gathers. Now the layered media reflectors all go away and we have a gather containing only two copies of the surface diffraction.

To study the surface scatter events, we need to be rid of the layered media primary events. These could be gotten rid of by the simple subtraction or they could be gotten rid of by a spatial lowpass filter or a spatial PEF, or in the event of gentle dip, by various kinds of "steering filters". We can return to the practical issue of separating the simple reflections from the surface scatter after we have covered some matters of principle.

2-D AUTOCORRELATION OF SURFACE SCATTERED RETURNS

Here we show theoretically that the 2-D autocorrelation (or 2-D spectrum) of surface scattered reflections is the same as that of the primary reflections. Thus by autocorrelation we will concentrate information that could be widely distributed in time and space. Later, we'll convert the autocorrelation to something more familiar.

Let the layered earth response from shot s to geophone g be $u(s, g, t) = u(0, g - s, t) = u(g - s, t)$ or in Fourier space, $u(g - s, \omega)$ or simply $u(g - s)$. When an upcoming wave hits the earth surface at g_1 it encounters a scattering object which reflects the primary wave with a random scaling $\xi(g_1)$. The signal at g_1 then takes off for a second flight like a multiple reflection, but departing in all directions. We are going to build the theoretical 2-D spectrum of this surface scattered wave w from the theoretical 2-D spectrum of u , the layered media primary reflection.

First we express the cascade of the two bounces. The arrival w at g_2 at time t is the sum of the time of each bounce, τ and $t - \tau$. Since this is a convolution in the time domain, we express it as a product in the frequency domain. Then we form the complex conjugate of this

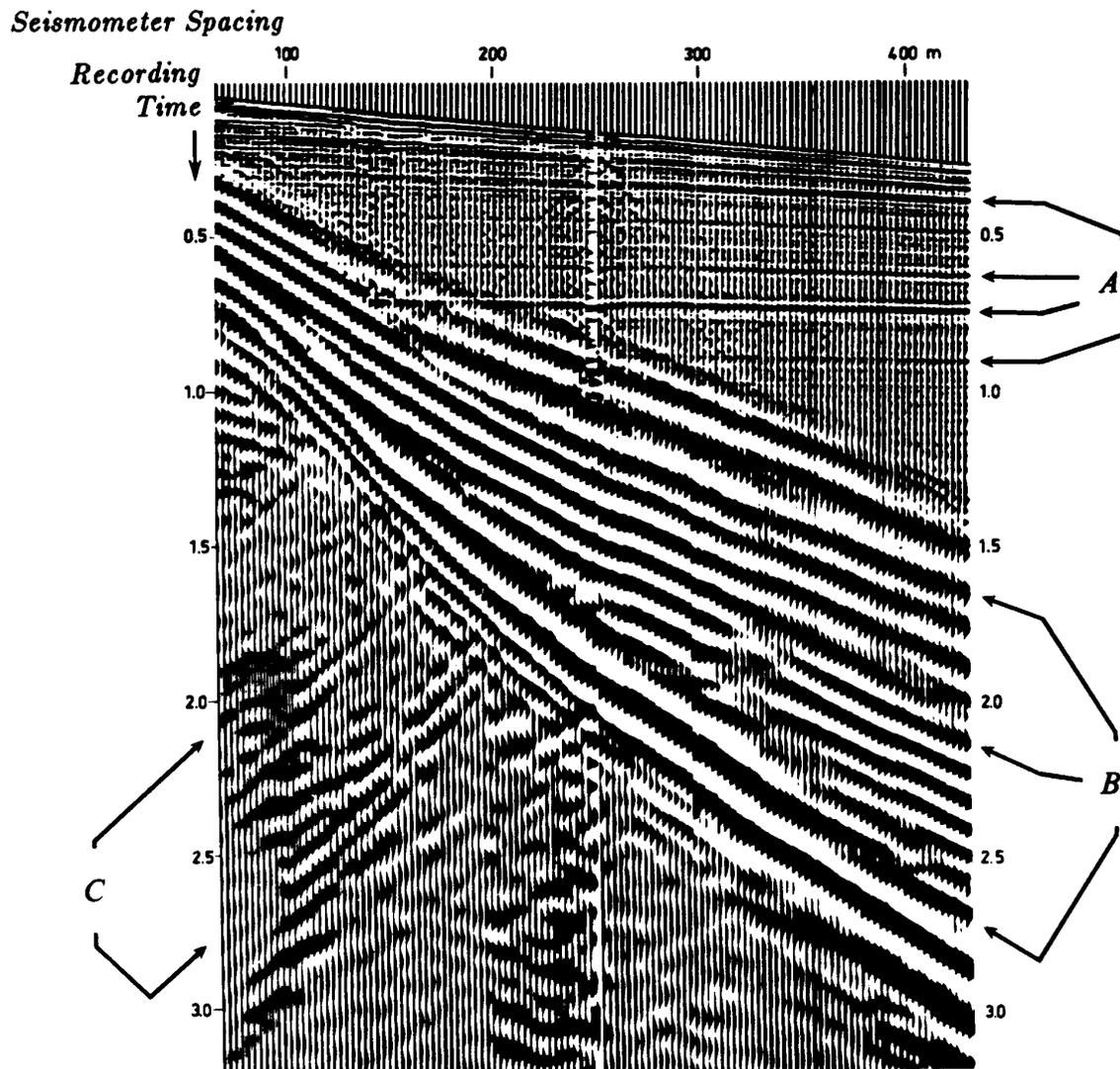


Figure 1: Shell shot gather from IEI page 172 (paper version) section 3.2. (/book/iei/fig/ofs/shell.ps) Observe the scattered ground roll with its apex at 250 m offset. It is interesting to identify the ray paths of ALL the backscattered energy. Might you believe the diffraction at 250 m offset is a weaker copy of the entire shot gather but shifted 250 m and .17 s? If you do, you will recognize that this copy also contains (albeit weakly) the near zero offset traces that are missing from the recording of the gather itself. jon4-shell [NR]

expression in preparation for autocorrelation on the x -axis.

$$w(s, g_2, t) = \sum_{g_1} \sum_{\tau} u(s, g_1, \tau) u(g_1, g_2, t - \tau) \xi(g_1) \quad (1)$$

$$w(s, g_2, \omega) = \sum_{g_1} u(g_1 - s, \omega) u(g_2 - g_1, \omega) \xi(g_1) \quad (2)$$

$$\bar{w}(s, g_2 + x, \omega) = \sum_{g_3} \bar{u}(g_3 - s, \omega) \bar{u}(g_2 + x - g_3, \omega) \bar{\xi}(g_3) \quad (3)$$

We insert the last two expressions into the expression for spatial autocorrelation.

$$A(s, x) = \sum_{g_2} w(s, g_2) \bar{w}(s, g_2 + x) \quad (4)$$

We will determine $A(s, x)$ experimentally as described earlier. Here we will see its theoretical relation to the primary reflected field u .

$$A(s, x) = \sum_{g_1} \sum_{g_2} \sum_{g_3} u(g_1 - s) u(g_2 - g_1) \bar{u}(g_3 - s) \bar{u}(g_2 - g_3 + x) \xi(g_1) \bar{\xi}(g_3) \quad (5)$$

$$= \sum_{g_1} \sum_{g_2} \sum_{g_3} u(g_1 - s) u(g_2 - g_1) \bar{u}(g_3 - s) \bar{u}(g_2 - g_3 + x) \delta(g_1 - g_3) \quad (6)$$

$$= \sum_{g_1} \sum_{g_2} u(g_1 - s) u(g_2 - g_1) \bar{u}(g_1 - s) \bar{u}(g_2 - g_1 + x) \quad (7)$$

$$= \sum_{g_1} u(g_1 - s) \bar{u}(g_1 - s) \sum_{g_2} u(g_2 - g_1) \bar{u}(g_2 - g_1 + x) \quad (8)$$

$$= \left(\sum_h u(h) \bar{u}(h) \right) \left(\sum_h u(h) \bar{u}(h + x) \right) \quad (9)$$

We Fourier transform over x . The first factor above is not a function of space. It is merely a function of ω , say a filter $|f(\omega)|^2$. Thus our main result:

$$|w(s, \omega, k_x)|^2 = |f(\omega)|^2 |u(\omega, k_x)|^2 \quad (10)$$

We see that in principle, for each shot point s , we measure the spectrum of the impulse response of the layered medium.

Hazardous cross terms

Suppose that we had not gotten rid of the layered media terms. The observations would be $u + w$ and we would be autocorrelating that. Both u and w have essentially the same autocorrelation, but their cross-term is dangerous. It has a different form (as you can verify). This form multiplies the expectation of $\xi(g)$ which might be theoretically zero but in practice might often not be zero. In principle we eliminate the cross term by eliminating u from the observation $u + w$, but in practice this difficulty remains to haunt us wherever we find the earth is not sufficiently horizontally layered.

AUTOCORRELATION

Need for all autocorrelations to be the same

The surface diffracted events that we might see on a shot gather are probably all much the same. (Anyway, I am not prepared to discuss their differences. It seems to be an implication of the theory. This deserves a better explanation or more study.)

Assuming the diffractions are all the same, their autocorrelations are all the same. (Anyway, if not, we are in big trouble, because then we get a smearing together of all their different autocorrelations.) The autocorrelation combines all the diffractions despite their different locations. By combining them, it enhances them. The autocorrelation does the job of merging the energy of all the scatterers.

There is an issue of the cross correlation of one diffractor with another. We'll suppose the cross correlations cancel out because of the random superposition of many shifts and directions. In practice this could fail if there are a small number of very strong ones.

Several people suggested that I should investigate the effect of random scatterers spread throughout the earth instead of having them all at the surface. I agree that is an interesting model to study, but I feared it because we all recognize that reflectors at all different depths will produce different hyperbolas. We dare not autocorrelate such data until we have processed it so that all hyperbolas look the same.

Spectral factorization

The trouble with the autocorrelation of the CSG is that we are not accustomed to it. We don't know how to think about it. We would rather have the CSG itself. This suggests spectral factorization. In one dimension it is ancient knowledge that spectral factorization finds us an impulse response function of the system. Using the helix (Claerbout, 1998), in helioseismology (Rickett and Claerbout, 1999a) (Rickett and Claerbout, 1999b) we found that we can recover a multidimensional causal acoustic impulse response of the sun instead of "autocorrelation wavelets" that helioseismologists had been getting. Now we conjecture that we can do something similar here:

CONJECTURE: A spectral factorization of the (autocorrelation of the) shot-geophone reflection data should give us the "multidimensional impulse response" of the earth.

There are some pitfalls with spectral factorization, but they are fairly well understood and they have often been overcome in the past.

What if the autocorrelation is bigger than the cable?

Consider the case where the cable is much shorter than the spatial autocorrelation of the diffractor. This would appear to be a troublesome case. We may notice from the Shell gather in Figure 1 that a limited range of offsets could allow us to see all offsets because of the presence of the scatterers at all ranges. If, however, we measure an autocorrelation over a small range, could spectral factorization possibly construct a minimum phase wavelet over a large range? I fear not. On the other hand, we can enlarge the aperture of a single shot by the idea of synthetic aperture. Let us see how this might be done.

Data layout

Ordinarily we think of reflection data as three dimensional, $P(t, g, s)$. That is because we redefine time to begin anew at $t = 0$ for each shot. Now let us use the more natural time, the wall clock time during data acquisition. Suppose the gun fires every 10 s for 10,000 s. Thus we have 1000 shots along a horizontal survey line. At each receiver we have this entire 10,000 s signal. We have one such a signal at each geophone. There is no shot axis. Thus the data is intrinsically two-dimensional, $P(t, g)$. Next we use the helix, as always, to wrap both t and g into one super-long signal. Apply spectral factorization, and unwrap the helix. What we should have is an estimate of the simple CSG we began with.

What is new, however, is very new and very interesting. When data is autocorrelated, it is averaged. In any average, some of the terms may be omitted if the sum is normalized properly.

I hypothesize that we'll have a very similar autocorrelation if we are missing many of the recordings. In particular, I propose to consider only the zero-offset traces. Forget about that long recording streamer! I hypothesize that the 2-D spectral factorization of the ZOS can give us a shot profile with all 1000 receivers.

CONJECTURE: The spectral factorization of the (autocorrelation of the) zero-offset section is the common midpoint gather.

This conjecture seems plausible when we recall that the ZOS amounts to the simple CSG convolved on the horizontal axis with a line of random numbers. The autocorrelation eliminates the random numbers and the spectral factorization recovers the CSG.

The proposal is really amazing: We could throw away our marine streamer and have only one receiver and hence only one point on the offset axis, yet the rocks randomly placed on the water bottom create for us a CMP gather that we could use for for velocity analysis. We better try it!

Finally, perhaps we can produce Cheops' pyramid.

CONJECTURE: The spectral factorization of the 2-D seismic data is Cheop's pyramid.

Spatial aliasing

An omnipresent theme in geophysical surveying is that we never have enough spatial coverage. In my book GEE, a theme is that we cope with insufficient data not via autocorrelation calculation but via prediction-error filter (PEF) estimation. Multidimensional PEFs are also a natural way to handle non stationarity. No doubt, we could return to the PEF approach for dealing with scattering too.

ACKNOWLEDGMENT

Professor Papanicolaou asserted that the superresolution phenomena is related to "sending a time reversed signal" back into the earth.

I note that if the impulse response of a system is time reversed and then fed into the same system, the output is an autocorrelation function. Many physical systems satisfy the conditions of reciprocity, that is, they have a symmetric matrix that carries source locations to receiver locations. On such a system you can interchange the sender and receiver and record the same signal.

My view here is that the magic comes from using autocorrelation. It can gather the scattered information that the main antenna is not prepared for.

I'd like also to thank James Rickett for his critical interest in this work, for helping me get my algebra correct, and for pointing out the danger with the cross term of $u + w$.

REFERENCES

- Claerbout, 1998, Multidimensional recursive filters via a helix: *Geophysics*, **63**, 1532–1541.
- Kolmogoroff, A., 1939, Sur l'interpolation et l'extrapolation des suites stationnaires: *C.R.Acad.Sci.(Paris)*, **208**, 2043–2045.
- Rickett, J., and Claerbout, J., 1999a, Acoustic daylight imaging via spectral factorization: *Helioseismology and reservoir monitoring: The Leading Edge*, **18**, 957–960.
- Rickett, J. E., and Claerbout, J. F., 1999b, Calculation of the acoustic solar impulse response by multi-dimensional spectral factorization: *Solar Physics*, accepted for publication.

Short Note

Common-azimuth migration and Kirchhoff migration for 3-D prestack imaging: A comparison on North Sea data

Louis Vaillant¹ and Henri Calandra²

INTRODUCTION

Common-azimuth migration (CAM) is a 3-D prestack depth migration technique based on the wave equation (Biondi and Palacharla, 1996). It exploits the intrinsic narrow-azimuth nature of marine data to reduce its dimensionality and thus manages to cut the computational cost of 3-D imaging significantly enough to compete with Kirchhoff methods. Based on a recursive extrapolation of the recorded wavefield, CAM is potentially better able to handle multi-pathing problems induced by complex velocity structures.

Elf Aquitaine provided us with an interesting dataset recorded in the North Sea, which shows a salt dome and other 3-D structures. The complexity of the wave propagation in the medium, resulting from high lateral and longitudinal velocity contrasts, yields multi-pathing and illumination problems. Vaillant and Sava (1999) have already illustrated how this model is both a serious challenge for imaging and an interesting test case for the CAM method.

In this paper, we present our latest imaging results that cover the whole vertical extent of the salt body with the CAM technique, as a complementary study to Vaillant and Sava (1999). At the same time, we compare the CAM-migrated cube to the image obtained using a state-of-the-art Kirchhoff algorithm and discuss the specificities of both approaches.

PREPROCESSING AND COMMON-AZIMUTH MIGRATION

The wave-equation approach obliges us to transform the 3-D data acquired with complex irregular geometry to regularly space-sampled data because CAM operates in the frequency-wavenumber domain and requires Fourier transforms along all axes. In contrast, Kirchhoff algorithms can handle irregular geometries without such preprocessing because they operate sequentially in a trace-by-trace manner.

¹email: louis@sep.stanford.edu

²Elf Exploration Production, Pau, France
email: henri.calandra@elf-p.fr

Here, data regularization is performed by an operator called AMO or Azimuth Moveout (Biondi et al., 1998). It also allows a local coherent stack that will reduce data volume. AMO, however, has a non-negligible computational cost in the whole imaging process (about 10% of migration cost). Effectively, introduced as the cascade of DMO and inverse DMO, AMO is a partial migration operator. With less accuracy, one can instead use a simple normalized binning procedure.

The processing scheme begins by creating a 5-axis time-midpoint-offset grid (t, \vec{m}, \vec{h}) for the data volume. Then, we apply a simple sequence NMO/AMO/NMO⁻¹ to regularize the data on the grid. This gridding procedure concurrently allows data resampling in common-midpoint and offset at the limit of aliasing, thus reducing further the cube dimensions and lower migration computational cost.

Marine data are usually concentrated within a narrow range of azimuth, as opposed to land data. Here, besides regularizing data along space axes, we use AMO to sum data coherently over the cross-line offset axis h_y ; conventionally, the subscripts x and y refer to the in-line and the cross-line direction, respectively. Thus, we obtain 4-D common-azimuth data, for which $h_y = 0$. After transformation to the frequency-wavenumber domain, this 4-D common-azimuth regularized dataset $D(\omega, \vec{k}_m, k_{hx})$ is the wavefield recorded at depth $z = 0$, to which CAM is applied.

Migration is then performed iteratively through common-azimuth downward-continuation of the wavefield (Biondi and Palacharla, 1996). The common-azimuth downward-continuation operator is derived from the stationary-phase approximation of the full 3-D prestack downward continuation operator. For more accuracy with lateral velocity variations, we use several reference velocities and interpolation as in the extended split-step method (Stoffa et al., 1990). The following chart summarizes the preprocessing and imaging schemes:

$$\begin{aligned}
 \tilde{D}_{irr}(t) &= \text{Gridding} \Rightarrow \tilde{D}(t, \vec{m}, \vec{h}) \\
 \tilde{D}(t, \vec{m}, \vec{h}) &= \text{AMO} \Rightarrow D(t, \vec{m}, h_x) \\
 D_{z=0}(\omega, \vec{k}_m, k_{hx}) &= \text{Down} \Rightarrow D_z(\omega, \vec{k}_m, k_{hx}) \\
 D_z(\omega, \vec{k}_m, k_{hx}) &= \text{Imaging} \Rightarrow D_z(\tau = 0, \vec{m}, h_x = 0)
 \end{aligned}$$

In this chart, \tilde{D} refers to data as an irregular set of traces, whereas D means data are a regular n-D cube.

APPLICATION TO REAL DATA

The migrated volume covers an area of 10.5×4 km, down to 5km. The velocity model, courtesy of Elf Aquitaine, was obtained by reflection tomography using the SMART method (Jacobs et al., 1992). For migration, we used 6 reference velocities and interpolation in an extended split-step manner. With data dimensions indicated in Table 1, the 3-D prestack common azimuth migration ran in 40 days on 4 processors of our SGI Power Challenge (18 MIPS R8000 processors).

Cmp-X	Cmp-Y	Depth	Offset	Frequencies	Velocities
525	160	400	64	176	6

Figure 1 shows a typical set of sections in the middle of the migrated cube. In the in-line section, the shallower part (above 1500m) reveals high-frequency details accurately imaged: the migration enhanced a graben structure, with normal faults and rotated blocks, around location Cmp-X=8000m, close to the top. The horizontal section at depth 900m highlights complex patterns inside sedimentary layers, imaged with a high resolution, that can be interpreted as turbidite channels. The imaging of the salt body is satisfactory, with sediments clearly bent towards the salt flank on the left-hand side. The salt flank itself has not been well focused, but can be revealed by residual migration (Sava, 2000). Even with an imperfect velocity model, migration enhances deep layers and most of the bottom of the salt. Several layers below the salt are also well focused.

COMPARISON WITH KIRCHHOFF MIGRATION

Figures 3 and 4 show comparisons between CAM and Kirchhoff migration results. The Kirchhoff algorithm used is derived from a preserved-amplitude approach and selects the most energetic arrival. Both CAM and Kirchhoff migration use exactly the same velocity model.

Some of the most significant differences between both approaches are well-known: in the CAM algorithm, finite frequency wave propagation is modeled avoiding the asymptotic approximations necessary for Kirchhoff. Also, for deep targets, the migration cost increases as N_z^3 (number of depth samples) for Kirchhoff, whereas CAM cost only increases as N_z^2 . However, Kirchhoff methods allow target-oriented migrations where CAM has to perform downward-continuation of the whole wavefield.

Figure 4 shows that the in-line sections around the salt body are relatively comparable in quality. Globally, CAM seems to give better results at imaging sediments bending against the salt flank on the left-hand side. The most important differences are shown by the horizontal sections (Figure 3): at a depth of 900m, CAM enhances complex high-frequency turbiditic patterns in shallow layers. At the same location, the Kirchhoff image appears at a considerably lower frequency and blurred along the in-line direction.

There are potentially 3 factors that could explain CAM's better accuracy compared to Kirchhoff migration:

- Wave propagation is handled completely differently: CAM iteratively propagates the wavefield by regular depth steps, as opposed to Green functions for Kirchhoff that are evaluated by ray tracing, wavefront construction or other methods.
- Since Green functions are not calculated for every source but rather pre-computed on a coarser grid, Kirchhoff algorithms usually include critical interpolations of traveltimes

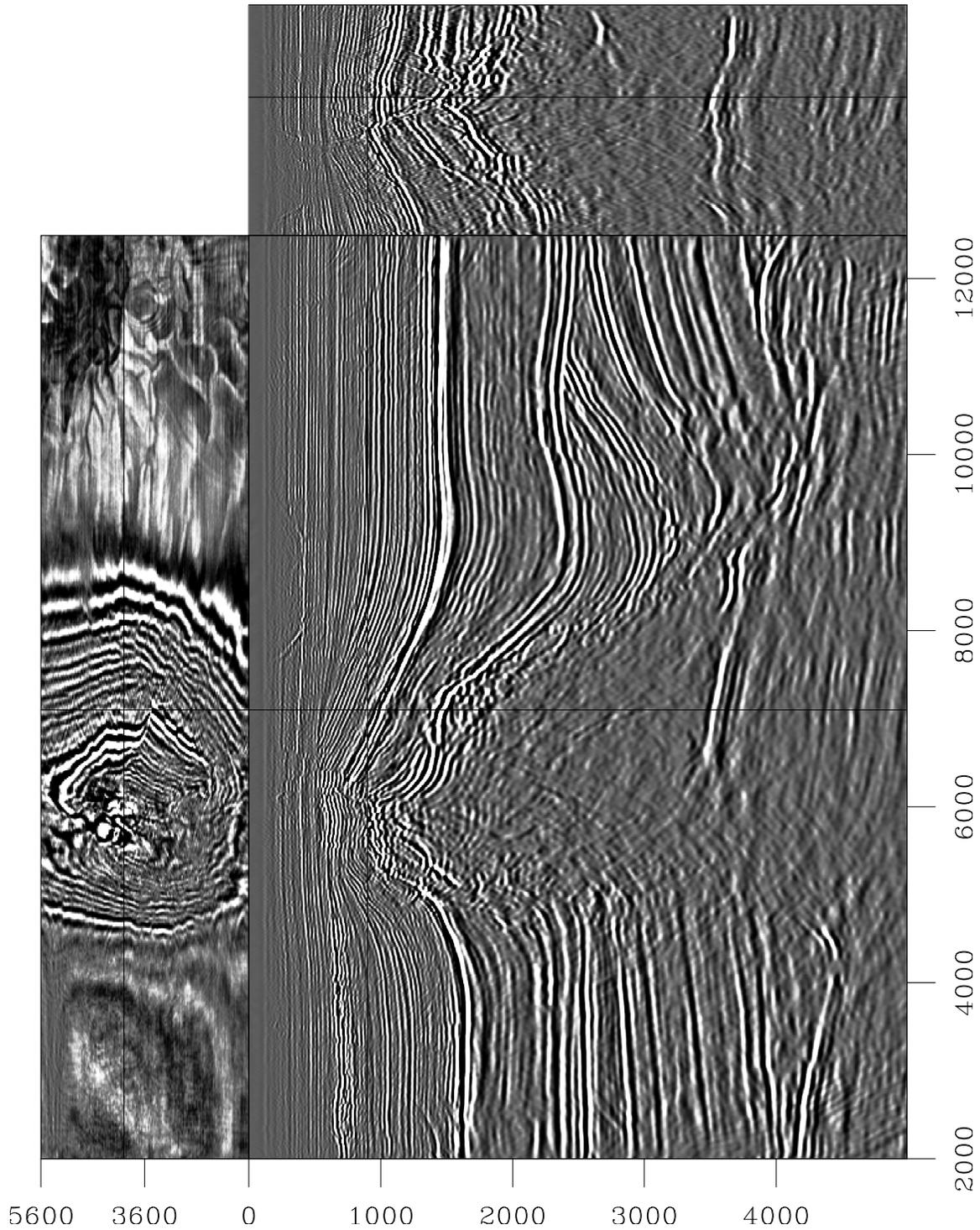


Figure 1: Sections of the image cube migrated with CAM: in-line $\text{Cmp-Y}=4000\text{m}$ (center), depth slice $z=900\text{m}$ (top), cross-line at $\text{Cmp-X}=7100\text{m}$ (side). `louis2-L7d-image-all-2` [CR]

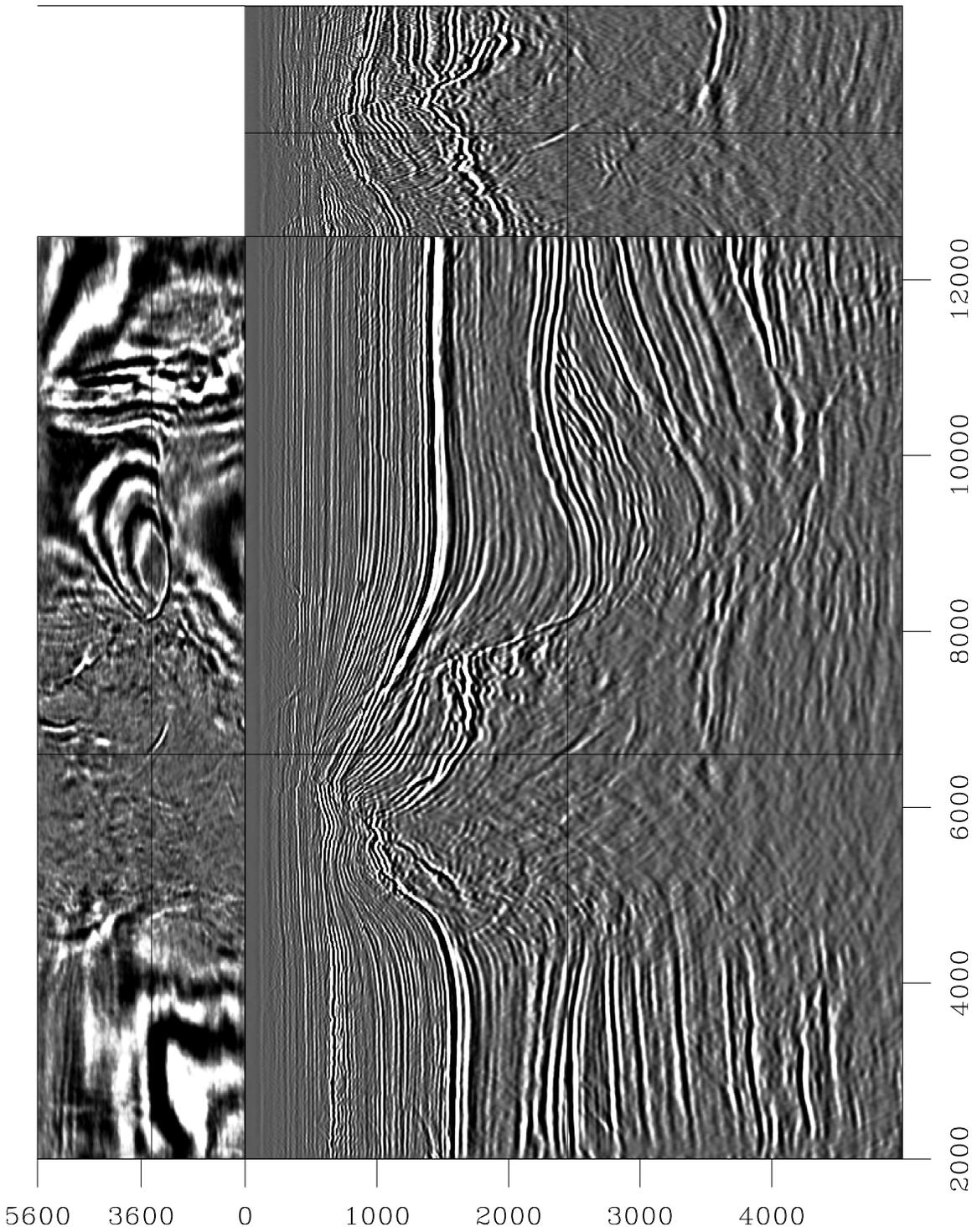


Figure 2: Sections of the image cube migrated with CAM: in-line Cmp-Y=3400m (center), depth slice z=2450m (top), cross-line at Cmp-X=6600m (side). louis2-L7d-image-all-2bis [CR]

maps and other attributes that may reduce accuracy for high-frequency details in the image. The only interpolations performed with CAM are the extended split-step scheme, which essentially addresses rapid velocity variations.

- Even if Kirchhoff algorithms can easily handle irregular geometries, the resulting image incorporates acquisition footprints, especially in depth slices. On the contrary, CAM requires regular geometry and acquisition problems are addressed during preprocessing through the AMO operator. Thus, the whole imaging stage is done with regularized geometry and potentially enables higher resolution in the final image.

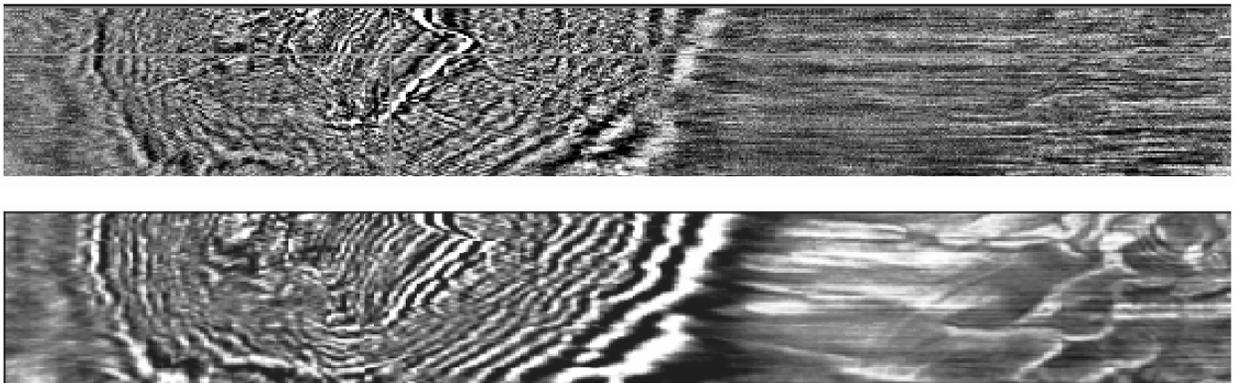


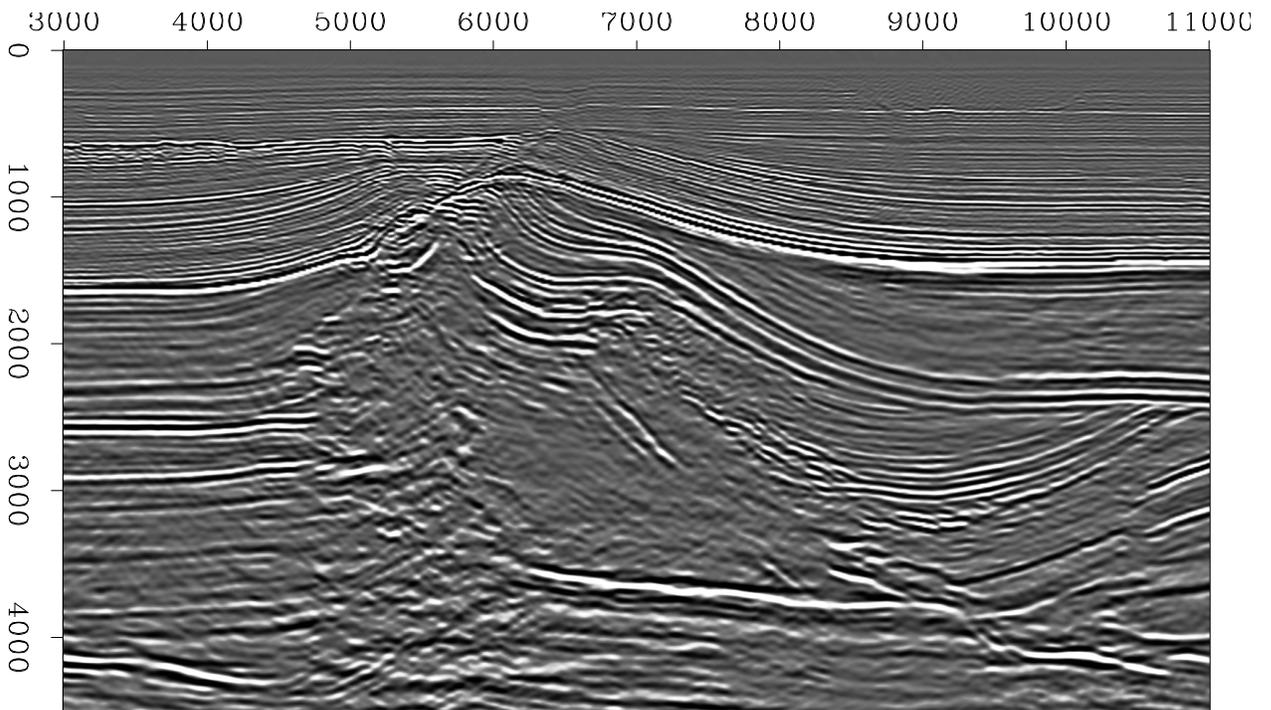
Figure 3: Comparison between Kirchhoff (top) and CAM (bottom) imaging results: depth slice at $z=900\text{m}$. The Kirchhoff image has a lower frequency content, a higher noise ratio and is blurred along the in-line direction. [louis2-kir-cam-zslice](#) [NR]

CONCLUSION

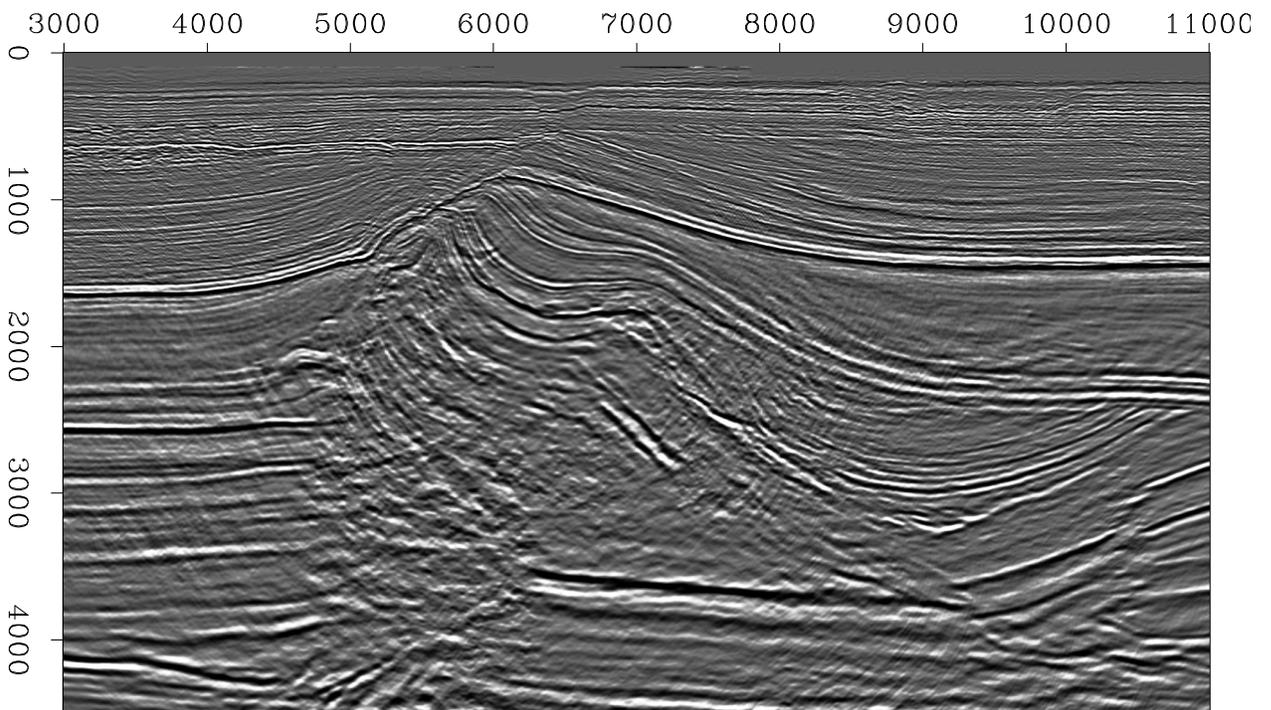
Real data offer an opportunity to test our imaging techniques further. Common-azimuth migration (CAM) is an attractive method for seismic imaging in complex media and it remains a subject for further research. Its computational cost and its high resolution in depth slices, illustrated in this particular example, can make it an attractive alternative to widespread Kirchhoff methods. Both approaches show different specificities and are equivalently valuable in our seismic imaging toolbox.

REFERENCES

- Biondi, B., and Palacharla, G., 1996, 3-D prestack migration of common-azimuth data: *Geophysics*, **61**, no. 6, 1822–1832.
- Biondi, B., Fomel, S., and Chemingui, N., 1998, Azimuth moveout for 3-D prestack imaging: *Geophysics*, **63**, no. 2, 574–588.



(a)



(b)

Figure 4: Comparison between CAM (a) and Kirchhoff (b) imaging results for the in-line section at Cmp-Y=3250m. `louis2-xz-CAM-KIR` [CR]

- Jacobs, J. A. C., Delprat-Jannaud, F., Ehinger, A., and Lailly, P., 1992, Sequential migration-aided reflection tomography: A tool for imaging complex structures: 62nd Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1054–1057.
- Sava, P., 2000, Variable-velocity prestack Stolt residual migration with application to a North Sea dataset: SEP-103, 147–157.
- Stoffa, P. L., Fokkema, J. T., de Luna Freire, R. M., and Kessinger, W. P., 1990, Split-step Fourier migration: Geophysics, **55**, no. 4, 410–421.
- Vaillant, L., and Sava, P., 1999, Common-azimuth migration of a North Sea dataset: SEP-102, 1–14.

Variable-velocity prestack Stolt residual migration with application to a North Sea dataset

Paul Sava¹

ABSTRACT

This paper investigates the applicability of prestack Stolt residual migration when the original image is obtained using an arbitrary velocity model. At its origin, the method is based on an assumption of constant velocity. However, its formulation for depth-migrated images involves a ratio of the reference and target velocities; therefore, for residual migration it is completely irrelevant if the original migration uses constant or variable velocity. Several examples, both on synthetic and real data demonstrate the effectiveness of the method. When applied to a North Sea dataset, the method highlights important features not seen in the original migration.

INTRODUCTION

Prestack Stolt residual migration has recently gained a new momentum because of its usefulness in imaging and migration velocity analysis (Biondi and Sava, 1999; Sava and Biondi, 2000). The method's main merit is that instead of requiring an assumption about the magnitude of the target velocity model to which we want to residually migrate, it only calls for an assumption about the ratio of the reference velocity to the target velocity (Sava, 1999b,a). Since the method is, in essence, a Stolt-stretch technique, it inherits Stolt's speed and convenience.

As for other Stolt-type techniques, the derivation of the method is based on the assumption of constant velocity. However, if we start from a depth-migrated image, the residual migration process can be shown not to depend on the actual velocities anymore, but rather on the ratio of the reference (v_0) and correct (v) velocities.

Strictly speaking, the ratio corresponds to two constant velocities. This would imply, however, that we obtain the original image using constant velocity as well, which is normally not the case. In this paper, I investigate the applicability of the residual migration method to images obtained using variable velocity models. I begin with a brief review of the theory and continue with three synthetic examples and one real-data example, which demonstrate that, although not theoretically accurate, the method can still be used in cases of variable velocity media.

¹email: paul@sep.stanford.edu

THEORY REVIEW

Stolt residual migration is formulated both for post-stack and prestack data. Stolt (1996) shows that, for the post-stack case, the residual migration operator is a function of the difference between the squared values of the original velocity and the new velocity to which the image is residually migrated. However, Stolt shows that such a formulation is no longer possible in the prestack case.

Sava (1999b,a) reformulates Stolt residual migration for depth-migrated images and shows that the operator is a function of the ratio of the two velocities, before and after residual migration, a formulation that holds for both post-stack and prestack (2-D, 3-D and common-azimuth) data.

In the 3-D case, the residual migration equation takes the form

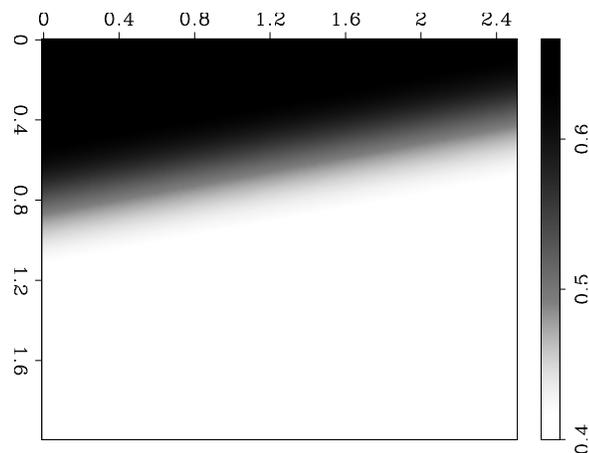
$$k_z = \frac{1}{2} \sqrt{\frac{v_0^2}{v^2} \frac{\left[4k_{z_0}^2 + \left(|\vec{k}_m + \vec{k}_h| - |\vec{k}_m - \vec{k}_h|\right)^2\right] \left[4k_{z_0}^2 + \left(|\vec{k}_m + \vec{k}_h| + |\vec{k}_m - \vec{k}_h|\right)^2\right]}{16k_{z_0}^2}} - \left|\vec{k}_m - \vec{k}_h\right|^2} + \frac{1}{2} \sqrt{\frac{v_0^2}{v^2} \frac{\left[4k_{z_0}^2 + \left(|\vec{k}_m + \vec{k}_h| - |\vec{k}_m - \vec{k}_h|\right)^2\right] \left[4k_{z_0}^2 + \left(|\vec{k}_m + \vec{k}_h| + |\vec{k}_m - \vec{k}_h|\right)^2\right]}{16k_{z_0}^2}} - \left|\vec{k}_m + \vec{k}_h\right|^2}, \quad (1)$$

where

- k_{z_0} and k_z are depth wavenumbers before and after residual migration,
- v_0 and v are, respectively, the reference and the target migration velocities, and
- \vec{k}_m and \vec{k}_h are the offset and midpoint wavenumbers.

Figure 1: The slowness map used for the first 2-D synthetic examples.

`paul1-dipsynt.slo` [CR]



Strictly speaking, Equation (1) is derived from the constant velocity dispersion relations for pre-stack data. However, since the new images do not depend on the true magnitude of the new velocity, but rather on its ratio to the original, we can still technically apply residual migration, even when the original velocity is variable. The question is how valid such a residual migration operation is for the case of variable velocity, and whether the output provides

any valuable information. A theoretical answer is not straightforward, since most likely the answer is data-dependent. I will, therefore, address the problem in an experimental way, first on synthetic models, and then on a North Sea real dataset.

SYNTHETIC EXAMPLES

I begin with two simple synthetic models. The first consists of a succession of flat reflectors embedded in a variable velocity model (Figure 1).

The left panel of Figure 2 shows the result of migration with the correct velocity, while the middle panel shows the result of migration with a scaled version of the original, and the right panel shows the result of residual migration using the correct velocity ratio. The flat reflectors are all restored to their original position, and the angle-domain common-image gathers are flat, indicating correct migration.

Figure 3 depicts a more complex model. In this case, the reflectors have different dips, and the velocity is variable and identical to the one in the preceding case (Figure 1). Again, the left panel shows the migration result with the correct velocity, the middle panel the result of migration with an incorrect velocity, and the right panel the result of residual migration with the correct velocity ratio. The reflectors are restored to their correct position and the angle-domain common-image gathers are flat.

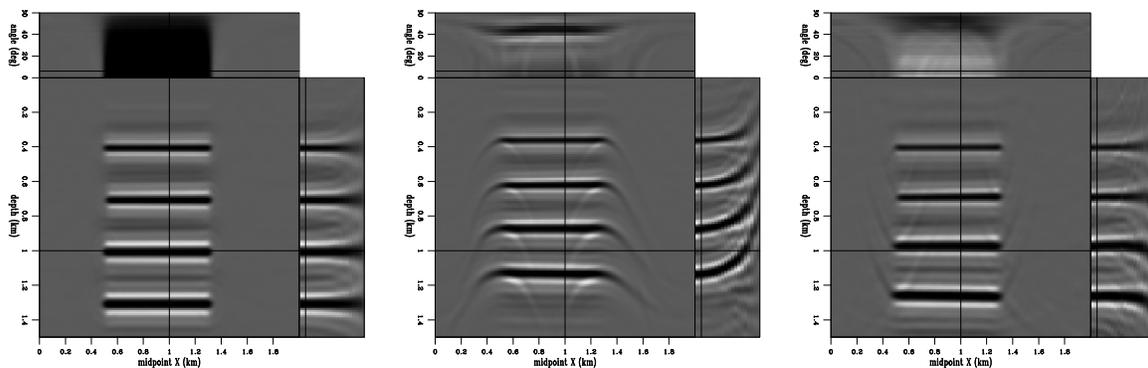


Figure 2: A simple synthetic model made of flat reflectors. The left panel represents migration with the correct velocity. The middle panel represents migration with an incorrect velocity, and the right panel represents the result of residual migration applied to the image in the middle panel. The reflectors are restored very close to their correct position. `paul1-flatsynt.cig` [CR]

The next synthetic example is represented by an image around a salt body. As in the preceding examples, the left panel of Figure 4 shows the result of migration with the correct velocity, in the middle panel the result of migration with an incorrect velocity, and in the right panel, the result of residual migration with the correct velocity ratio. Again, we can conclude that prestack Stolt residual migration is able to recover the correct image, even when the velocity map is not constant. We can, therefore, apply the procedure to a much more complex seismic image obtained from data recorded over a North Sea salt dome.

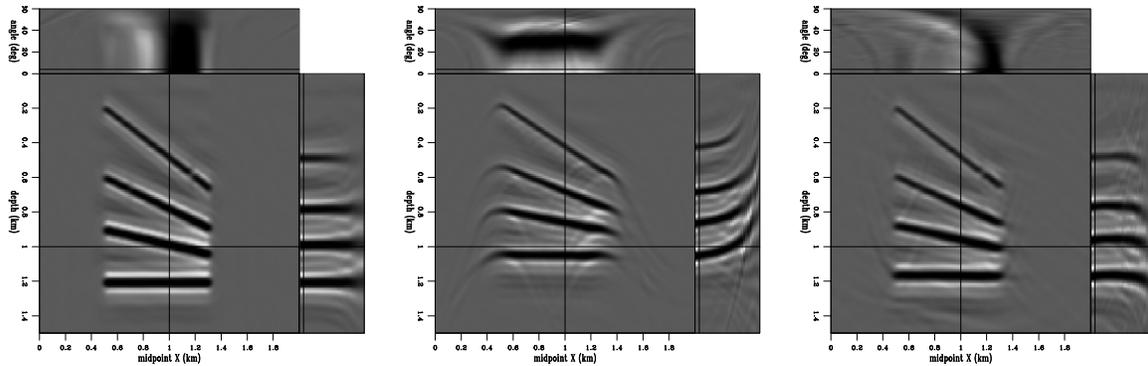


Figure 3: A simple synthetic model made of dipping reflectors. The left panel represents migration with the correct velocity. The middle panel represents migration with an incorrect velocity, and the right panel represents the result of residual migration applied to the image in the middle panel. The reflectors are restored very close to their correct position. `paul1-dipsynt.cig` [CR]

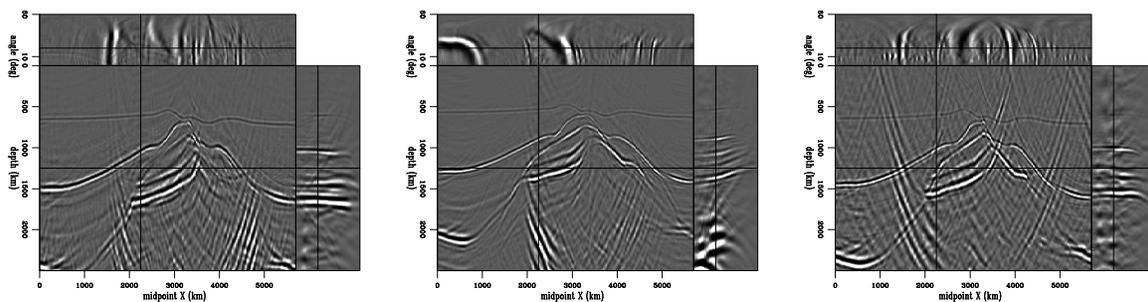


Figure 4: A complex synthetic model around a salt body. The left panel represents migration with the correct velocity. The middle panel represents migration with an incorrect velocity, and the right panel represents the result of residual migration applied to the image in the middle panel. The reflectors are restored to their correct position. `paul1-saltsynt.cig` [CR]

A NORTH SEA EXAMPLE

The North Sea dataset used in this paper has been previously discussed by Vaillant and Sava (1999). Figure 5 shows a 2-D slice out of the image-cube obtained by common-azimuth migration (Biondi and Palacharla, 1996). Without going into many details, we can observe that many of the important features of the image have been well resolved, especially away from the salt. However, there still are some regions in the image insufficiently clarified, as labeled in Figure 5:

1. **(A)** Where is the salt overhang flank?
2. **(B)** How do the sediments terminate against the salt?
3. **(C)(D)** What is this?

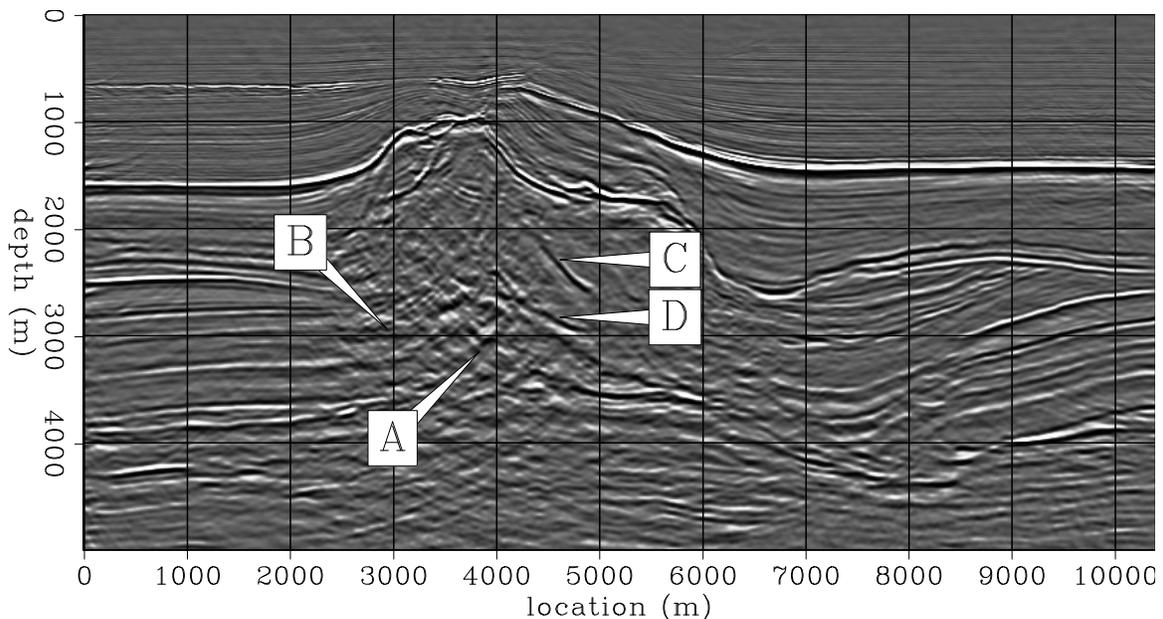
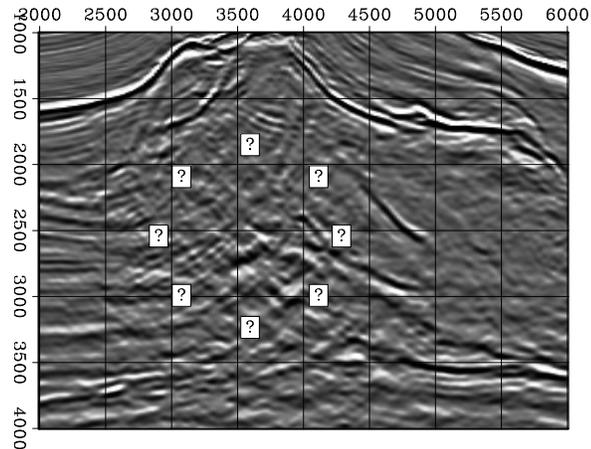


Figure 5: Seismic image for the North Sea data. Major questions still unanswered are **A** Where is the salt overhang flank? **B** How do the sediments terminate against the salt? **C** **D** What is this? [paul1-saltreal.problem](#) [CR]

Essentially, an important area of the image, close to the overhang, remains under severe blur caused by the salt body. This area, encircled with question marks in Figure 6, is the focus for the residual migration processing. Two hypotheses can be made about the cause of the blur:

- it is caused by inaccuracies of the imaging algorithm, a hypothesis that is analyzed elsewhere (Vaillant and Calandra, 2000);

Figure 6: Zoom over the area of major blur. The goal of the residual migration is to shed as much light as possible over the region encircled with question marks. `paul1-saltreal.zoom` [CR]



- the velocity map around the salt body is not correct, a hypothesis which is the focus of this paper.

Since in this case we are analyzing a real dataset, we do not know what the correct velocity ratio ($\gamma = \frac{v_0}{v}$) is to which we would like to residually migrate. Most likely, in fact, there is no one velocity ratio that would solve the problem over the entire seismic image. What we need to do then is run residual migration for a range of velocity ratios and analyze what changes have occurred. In the end, we should either select one image which answers our questions, or interpolate an optimal image.

Figure 7 shows how the target image is changed when the parameter γ varies from 0.90 to 1.00. A value of $\gamma < 1.00$ indicates that we residually migrate the image with a *higher velocity* than the original. It is apparent that the overhang salt flank is significantly better imaged at higher velocities compared to the reference image. Also, many of the layers in the image continue under the salt, and therefore become much easier to interpret.

Figure 8 represents a face-to-face comparison of the original image (left panel) and a cleaner image obtained with residual migration (right panel), which enables us to answer some of the questions posed at the beginning of the discussion:

1. The salt flank (**F**) is more continuous and extends significantly more towards the surface. It is also more energetic, and positioned lower in the image. Since this reflector is imaged mostly with the salt velocity, we can conclude that the salt velocity is not high enough in the original model and therefore needs to be increased to achieve better imaging.
2. The reflector (**R**), previously imaged inside the salt, appears much fainter, perhaps simply indicating that this event is merely an artifact or a salt internal multiple.
3. The sediments below the salt **S** are significantly cleaner. What is practically not interpretable in the original image becomes much cleaner and makes a lot more geological sense. We can, in fact, trace how the sediments bend against the salt body.

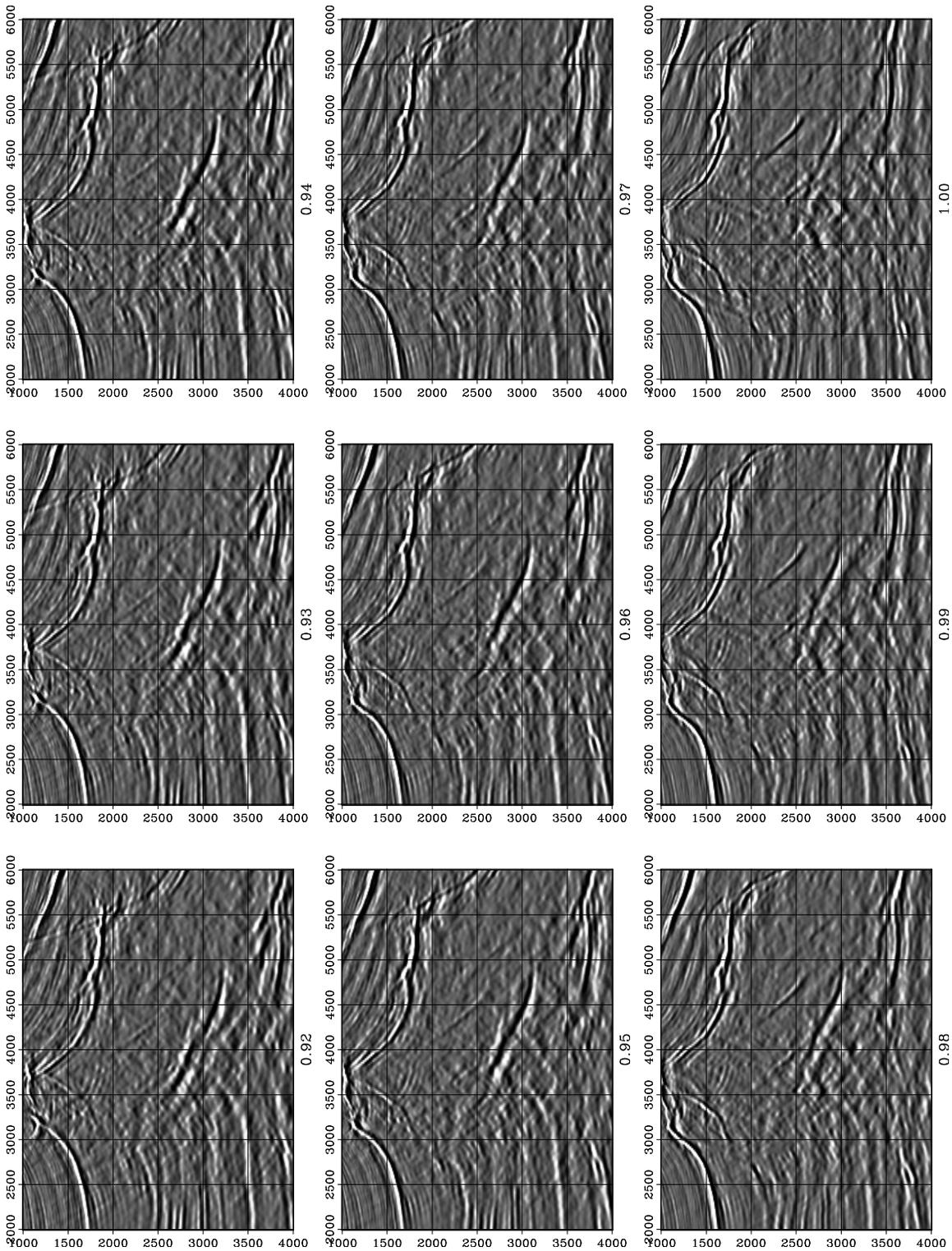


Figure 7: Residual migration results. Each panel represent the image corresponding to different values of the parameter γ , ranging from 0.92 to 1.00. The salt flank is significantly better imaged at lower values of γ , that is, at velocity values higher than those of the original model.

`paul1-saltreal.storm` [CR]

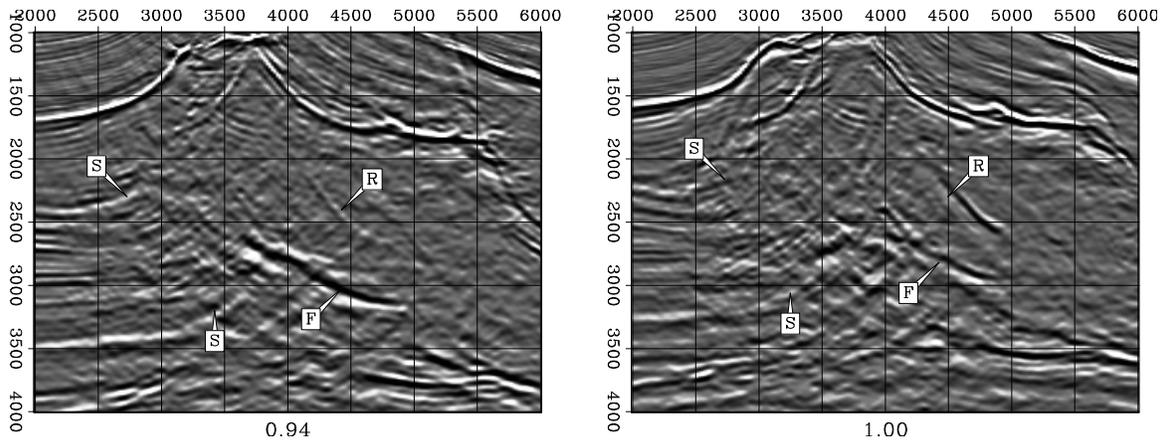


Figure 8: A face-to-face comparison of the original seismic image (left) and an image obtained by residual migration with the ratio $\gamma = 0.94$ (right). `paul1-saltreal.face2face` [CR]

The downside of selecting only an image corresponding to a constant parameter γ over the entire map is that the image degrades in the regions that do not require an increase in velocity, or that may only require a smaller increase. A good example is the upper part of the salt, which appears slightly over-migrated and defocused in the residual migration panel (Figure 8, left) with respect to the original image (Figure 8, right).

A reasonable solution to this problem is to pick different ratios at every location in the images in Figure 7. The interpolation aim is to preserve the original image in the regions where it is correct and to bring into the picture the regions that show improvement after residual migration. Figure 9 shows the result of optimal picking:

- The top part of the salt, between the horizontal locations 2500 and 3500 m, where the remaining diffraction in the original image is much better collapsed at a higher velocity ($\gamma = 0.97$).
- The overhang salt flank and the sediments underneath are much cleaner after residual migration than in the original. This situation also corresponds to a higher velocity ($\gamma = 0.94$).

CONCLUSIONS

This paper shows that prestack Stolt residual migration, in spite of its constant velocity origin, can be successfully applied to depth images obtained with an arbitrary velocity. Several examples, both on synthetic and on real data, prove the effectiveness of the method. The key to improving the seismic images lies in correctly picking the variable surface of the parameter γ , rather than in selecting only one image at a particular value.

The method also shows that the blurred regions of the seismic image of the North Sea dataset are not so much an imaging problem as a velocity problem: a better velocity map

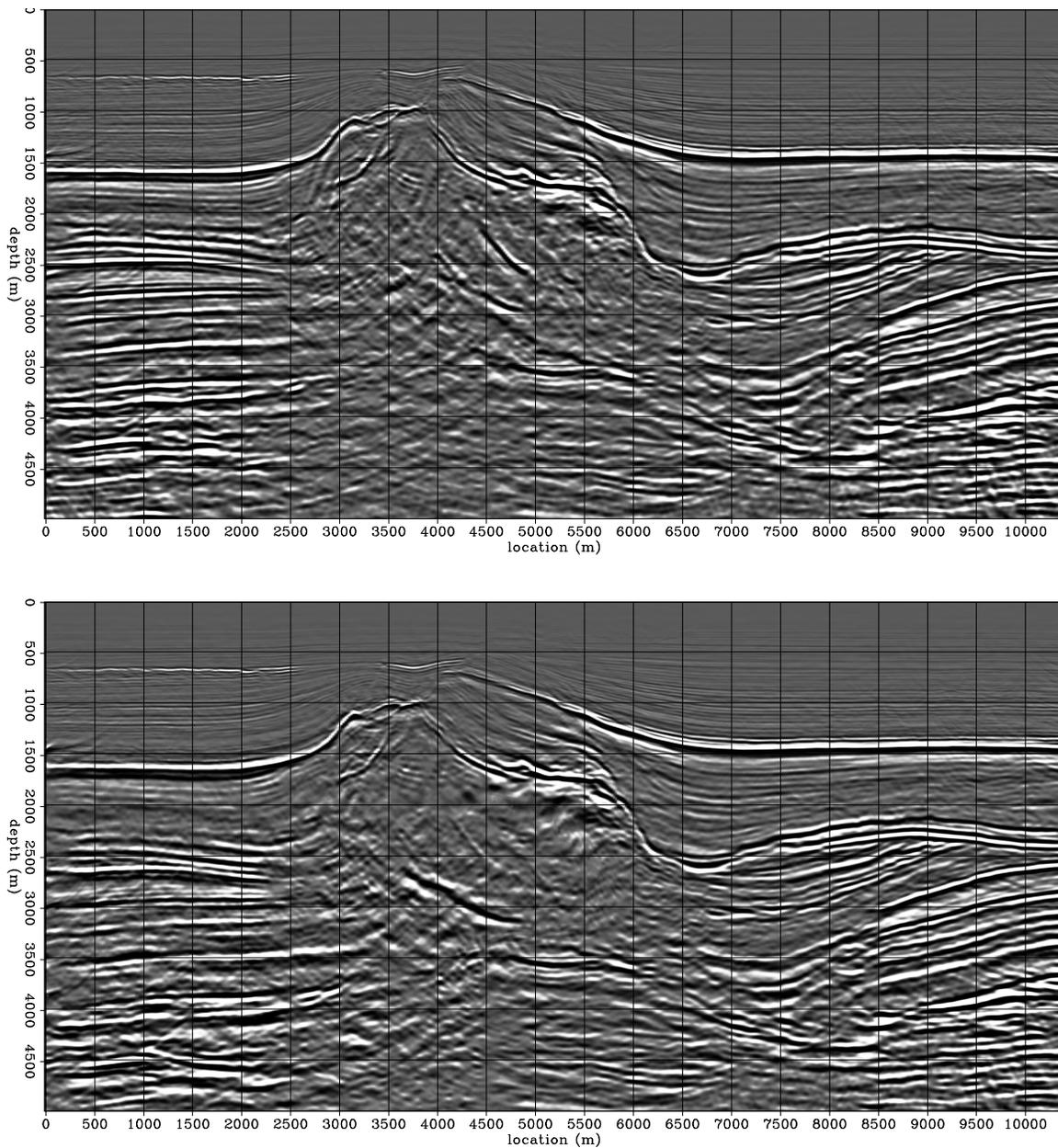


Figure 9: Comparison between the images before and after residual migration. The top panel is the image obtained with common-azimuth migration on the original velocity model. It shows incomplete focusing in the upper part of the salt and blurred salt overhang. The bottom panel is the image after residual migration, and shows a collapsed diffraction at the top of the salt, and much better imaged salt overhang and under-salt sediments. paul1-saltreal.comparison [CR]

could enable a good algorithm to image the overhang of the salt, at least partially. Prestack Stolt residual migration is able to improve the migrated image, at a fraction of the cost of the original migration.

ACKNOWLEDGMENTS

I would like to thank Biondo Biondi for many stimulating discussions. Elf Aquitaine provided both the synthetic salt-dome model and the North Sea dataset.

REFERENCES

- Biondi, B., and Palacharla, G., 1996, 3-D prestack migration of common-azimuth data: *Geophysics*, **61**, no. 6, 1822–1832.
- Biondi, B., and Sava, P., 1999, Wave-equation migration velocity analysis: *SEP-100*, 11–34.
- Sava, P., and Biondi, B., 2000, Wave-equation migration velocity analysis: Episode II: *SEP-103*, 19–47.
- Sava, P., 1999a, Short note—on Stolt common-azimuth residual migration: *SEP-102*, 61–66.
- Sava, P., 1999b, Short note—on Stolt prestack residual migration: *SEP-100*, 151–158.
- Stolt, R. H., 1996, Short note—a prestack residual time migration operator: *Geophysics*, **61**, no. 2, 605–607.
- Vaillant, L., and Calandra, H., 2000, Common-azimuth migration and Kirchhoff migration for 3-D prestack imaging: A comparison on North Sea data: *SEP-103*, 139–147.
- Vaillant, L., and Sava, P., 1999, Common-azimuth migration of a North Sea dataset: *SEP-102*, 1–14.

Accuracy of common-azimuth migration approximations

*Louis Vaillant and Biondo Biondi*¹

ABSTRACT

Common-azimuth migration (CAM) is an attractive solution for 3-D prestack imaging. It reduces the full 5-D phase-shift operator to 4-D through the stationary-phase approximation, lowering the computational cost. However, this assumption yields constraints in the downward-continuation process that can limit accuracy. Those errors are estimated in this paper by comparison to other wave-equation methods and to analytical solutions. Common-azimuth migration appears robust, but leaves opportunities for formulating an extended migration algorithm, which overcomes some of its inherent limits.

INTRODUCTION

Wave-equation migration techniques offer an attractive alternative to widespread Kirchhoff methods for 3-D prestack depth migration, with modern powerful computing resources. Several authors have recently illustrated these techniques' potential for handling multi-pathing problems and complex velocity media (??). This study focuses on the particular case of the common-azimuth migration (CAM) method and discusses the accuracy of its approximations.

Common-azimuth migration is a 3-D prestack depth migration technique based on the wave equation (?). It exploits the intrinsic narrow-azimuth nature of marine data to reduce its dimensionality. Migration is performed iteratively through common-azimuth downward-continuation of the wavefield. This common-azimuth downward-continuation operator is derived from the stationary-phase approximation of the full 3-D prestack downward continuation operator. Thus, the CAM approach manages to cut the computational cost of 3-D imaging significantly enough to compete with Kirchhoff methods.

Even though CAM is designed for 3-D migration in complex media, we used here only synthetic data and 1-D velocity models. Our purpose was to identify better its behavior on simple examples. In this paper, we first discuss wave propagation in constant gradient velocity media and then analyze migration results of synthetic data.

¹email: louis@sep.stanford.edu, biondo@sep.stanford.edu

RAYS IN CONSTANT GRADIENT VELOCITY MEDIA

Analytical ray tracing

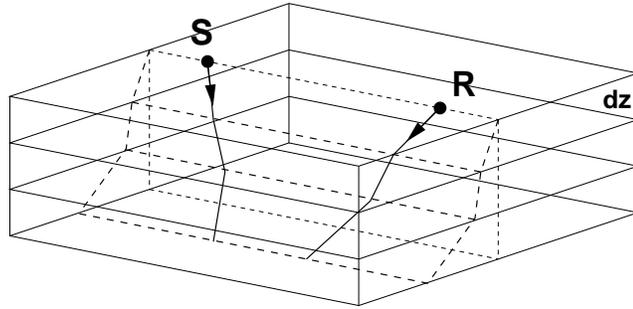
Theory shows that even in a simple 1-D $v(z)$ medium, CAM is not perfectly accurate: the stationary-phase approximation used in the derivation of CAM (?) imposes the relation among ray parameters indicated below that constrains rays to keep the same azimuth at each depth step (Figure 1):

$$\frac{p_{sy}}{p_{sz}} = \frac{p_{ry}}{p_{rz}}, \quad (1)$$

where the subscripts s and r refer to the rays coming from the source and the receiver, respectively.

Figure 1: Ray geometry imposed by common-azimuth constraints: both receiver and source rays keep the same azimuth at each depth step.

`louis1-ray-comaz` [NR]



We choose to test the behavior of rays and the accuracy of CAM approximations in a synthetic medium where the velocity varies linearly. In such a medium, ray trajectories can be computed analytically, as well as all ray parameters. Figure 2 illustrates the geometry of the problem. With those notations, ray curvature can be expressed as (?):

$$\kappa = \frac{\|\nabla \mathbf{v}\|}{v} \sin \theta \quad (2)$$

In constant gradient velocity media where $v(z) = v_0 + \gamma z$, the ray curvature is thus constant, i.e., rays are portions of circles included in a vertical plane. The radius of those circles is

$$R = \frac{1}{\kappa} = \frac{v(z)}{\gamma \sin \theta(z)} = \frac{v_0}{\gamma \sin \theta_s} \quad (3)$$

The ratio $\frac{\sin \theta(z)}{v(z)}$ is also the horizontal component p_ρ of the slowness vector along the ray, which therefore is also a constant:

$$R = \frac{1}{\kappa} = \frac{1}{\gamma p_\rho} = \text{cst} \quad (4)$$

After some calculation (see Appendix), the equation of the circle of radius R passing through point source $S(\rho_s, z_s)$ with initial incident angle θ_s is, in the plane (S, ρ, z) :

$$\left(\rho - \rho_s - \frac{\cos \theta_s}{\gamma p_\rho} \right)^2 + \left(z - z_s + \frac{v(z_s)}{\gamma} \right)^2 = \left(\frac{1}{\gamma p_\rho} \right)^2 \quad (5)$$

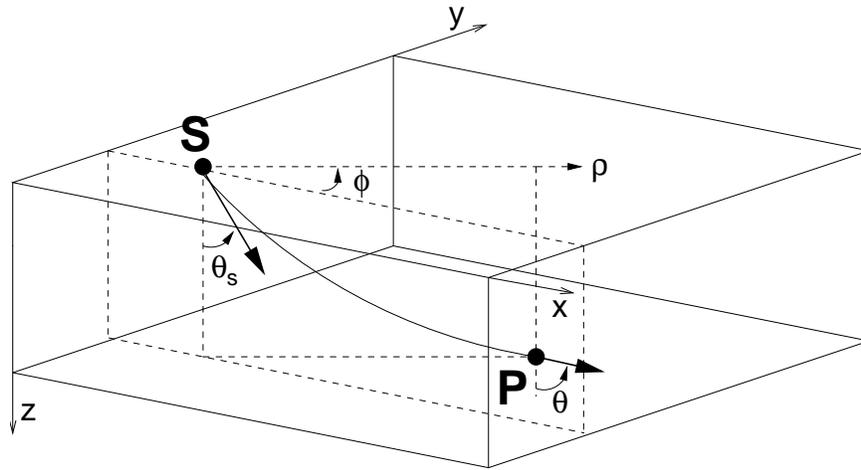
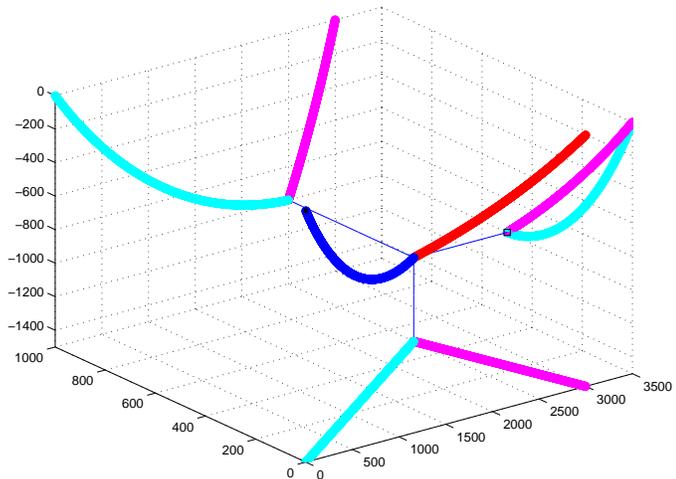


Figure 2: Ray geometry in a constant gradient velocity medium. `louis1-ray-vgrad` [NR]

For any given triplet of points (S,P,R), respectively source, image point and receiver locations, there exist only two circles satisfying equation (5) that form the complete ray path. Figures 3 to 5 illustrate such ray paths. We can verify that the projections of the source ray and the receiver ray on the cross-line plane do not coincide in general and therefore break the assumption of azimuth conservation in CAM downward-continuation imposed by relation (1).

Figure 3: Example of 3-D analytical ray tracing, with the three projections of both rays on vertical and horizontal planes. Source and receiver location are indicated with solid stars. The reflection point and its three projections are represented by a circle. Offset is 3000m (in-line). Velocity is $v(z) = 1500 + 0.5z$ m/s. `louis1-ray_exmpl1` [ER]



However, with too strong a velocity gradient, rays quickly start to overturn (Figure 4). The corresponding reflection cannot be imaged with one-way wave propagation methods, such as CAM and the other wave-equation migration methods we discuss in this paper.

Figure 4: Same geometry as before. The only difference is in the velocity: $v(z) = 1500 + z$ m/s. The source ray has overturned because of the stronger velocity gradient. `louis1-ray_exmpl2` [ER]

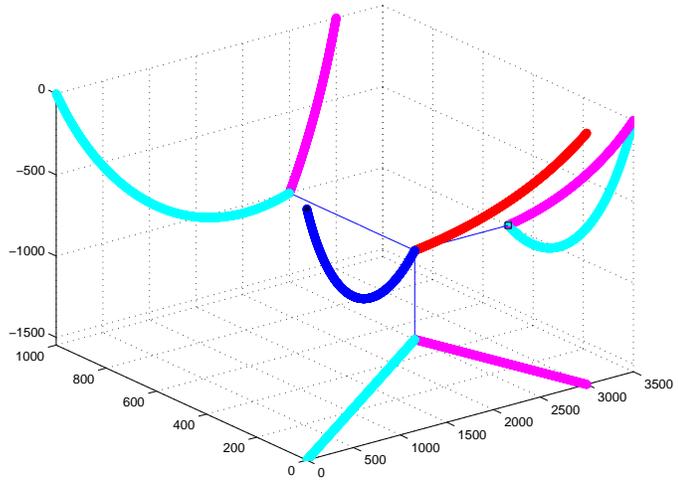
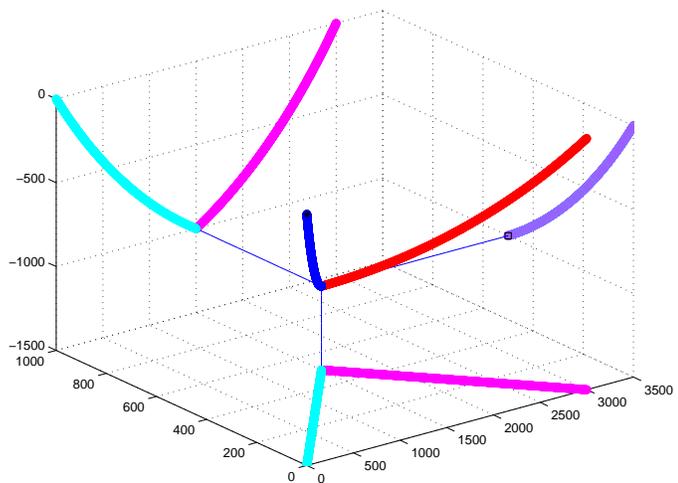


Figure 5: Velocity law is $v(z) = 1500 + 0.5z$ m/s. The reflection point is at an equal distance from source and receiver: the problem is symmetrical and azimuth is conserved at each depth step. `louis1-ray_exmpl3` [ER]



CAM stationary path

In the cartesian coordinate system, the components of the source and receiver slowness vectors along a ray are

$$p_{sx} = \frac{\sin \theta_s \cos \phi_s}{v(z)} \quad p_{rx} = \frac{\sin \theta_r \cos \phi_r}{v(z)} \quad (6)$$

$$p_{sy} = \frac{\sin \theta_s \sin \phi_s}{v(z)} \quad p_{ry} = \frac{\sin \theta_r \sin \phi_r}{v(z)} \quad (7)$$

$$p_{sz} = \frac{\cos \theta_s}{v(z)} \quad p_{rz} = \frac{\cos \theta_r}{v(z)} \quad (8)$$

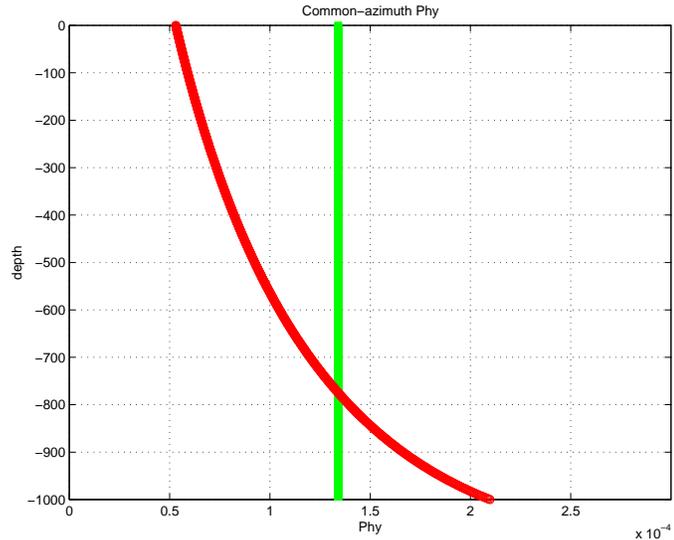
In this context, we can reformulate the expression for the stationary path in CAM theory (?), which gives the cross-line offset ray parameter as a function of velocity and ray parameters:

$$\hat{k}_{hy} = k_{my} \frac{\sqrt{1 - v(z)^2 p_{rx}^2} - \sqrt{1 - v(z)^2 p_{sx}^2}}{\sqrt{1 - v(z)^2 p_{rx}^2} + \sqrt{1 - v(z)^2 p_{sx}^2}} \quad (9)$$

Moreover, since wave propagation can be handled completely analytically in constant gradient velocity, we can calculate the theoretically “exact” cross-line offset wavenumber and compare it to the values given by the stationary-phase approximation (Equation (9)), as shown in Figure 6. Here, in the case of a reflection on a plane dipping at 60° and oriented at 45° with respect to the in-line direction, the stationary path given by Equation (9) is a seriously biased approximation.

Figure 6: Comparison of the exact cross-line offset ray parameter p_{hy} (thick solid line) and of the approximated p_{hy} (dashed curve) calculated with CAM stationary-phase approximation, in the case of a reflector dipping at about 60° and oriented at 45° with respect to the in-line direction.

`louis1-phy_cam` [ER]



MIGRATION OF SYNTHETIC DATA

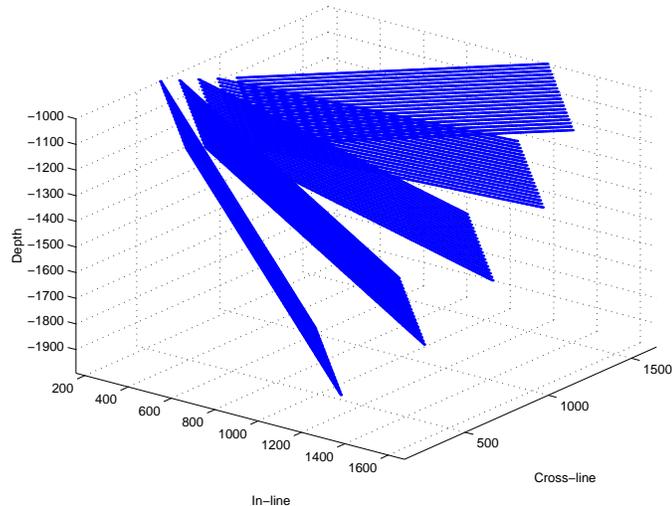
In order to test further the accuracy of CAM in constant gradient velocity media, we generated synthetic datasets and migrated them with the following wave-equation algorithms: Offset

plane-wave migration (?), CAM and full Phase-Shift (5-D operator).

Data modeling

Data were generated using SEPlib `kirmod3d` program. The reflectivity map simply consists of a set of gradually dipping planes, from zero dip to 60° . The azimuth of the planes is 45° with respect to the direction of the acquisition, which maximizes problems in imaging.

Figure 7: Geometry of the set of slanted planes, dipping at 0° , 15° , 30° , 45° and 60° towards increasing x and y , at 45° with respect to the in-line direction. `louis1-planes` [ER]



Data generated are on a regular grid and are common-azimuth, that is, with no cross-line offset component. The geometry of the grid is illustrated in Figure 8, with 64 samples in `cmp-x` and `cmp-y`, 128 in offset.

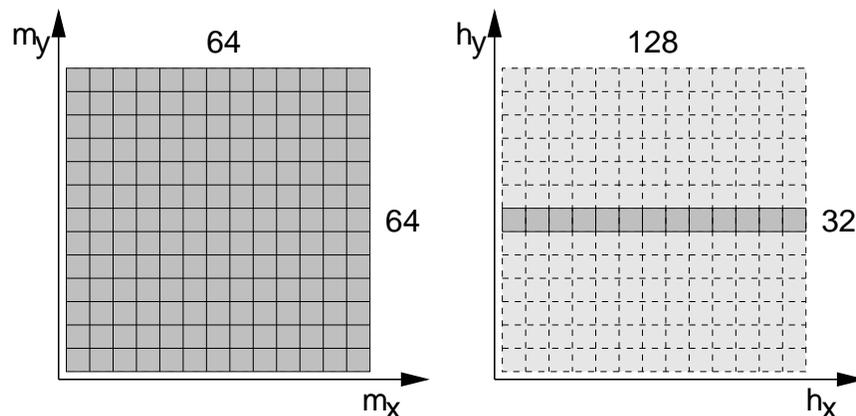


Figure 8: Geometry of the gridded synthetic data. The left represents a midpoint plane from the whole cube; the right is an offset plane. The common-azimuth cube has no cross-line offset, but can be zero-padded (light gray) in order to apply a 5-D phase-shift operator.

`louis1-dim-grid` [NR]

We used a very fine time sampling (1 ms) in order to minimize interpolation errors during

modeling with `kirmod3d` and sub-sampled data to 2ms afterwards. We generated Green functions with velocity law $v(z) = 1500 + 0.5z$, which roughly corresponds to typical gradients found in the Gulf of Mexico.

Examples with several wave-equation migration algorithms

Biondi and Vaillant (?) discuss the relative accuracy of offset plane-wave migration and CAM for wave-equation imaging. Both are derived from full downward continuation of 3-D prestack data with the Double Square Root (DSR) phase-shift operator:

$$k_z = \sqrt{\frac{\omega^2}{v^2(\mathbf{s}, z)} - \frac{1}{4} [(k_{mx} - k_{hx})^2 + (k_{my} - k_{hy})^2]} + \sqrt{\frac{\omega^2}{v^2(\mathbf{g}, z)} - \frac{1}{4} [(k_{mx} + k_{hx})^2 + (k_{my} + k_{hy})^2]} \quad (10)$$

where ω is the temporal frequency, \mathbf{k}_m is the midpoint wavenumber vector, \mathbf{k}_h is the offset wavenumber vector, $v(\mathbf{s}, z)$ and $v(\mathbf{g}, z)$ are the velocity at the source and receiver locations, respectively.

Offset plane wave (OPW) migration (?) performs migration of each offset plane wave component of the data independently. It can be interpreted as a reversed-order two-pass prestack migration, where an initial cross-line zero-offset migration is followed by an in-line prestack migration. Another interpretation is that the cross-line offset wavenumber k_{hy} in equation (10) is set to zero for downward continuation:

$$k_z = \sqrt{\frac{\omega^2}{v^2(\mathbf{s}, z)} - \frac{1}{4} [(k_{mx} - k_{hx})^2 + k_{my}^2]} + \sqrt{\frac{\omega^2}{v^2(\mathbf{g}, z)} - \frac{1}{4} [(k_{mx} + k_{hx})^2 + k_{my}^2]} \quad (11)$$

Instead, for CAM, the cross-line offset wavenumber k_{hy} is replaced in equation (10) by its stationary path \hat{k}_{hy} given in equation (9):

$$k_z = \sqrt{\frac{\omega^2}{v^2(\mathbf{s}, z)} - \frac{1}{4} [(k_{mx} - k_{hx})^2 + (k_{my} - \hat{k}_{hy})^2]} + \sqrt{\frac{\omega^2}{v^2(\mathbf{g}, z)} - \frac{1}{4} [(k_{mx} + k_{hx})^2 + (k_{my} + \hat{k}_{hy})^2]} \quad (12)$$

Both migrations reduce the full 5-D phase-shift operator to a 4-D operator. In fact, when no multipathing occurs, only a 4-D slice of the 5-D wavefield contributes to the image.

Figures 9 and 10 show migration results. The final image cube has 4 dimensions, the last being the common-image gather (CIG) ray parameter axis, generated by slant stack (?). The

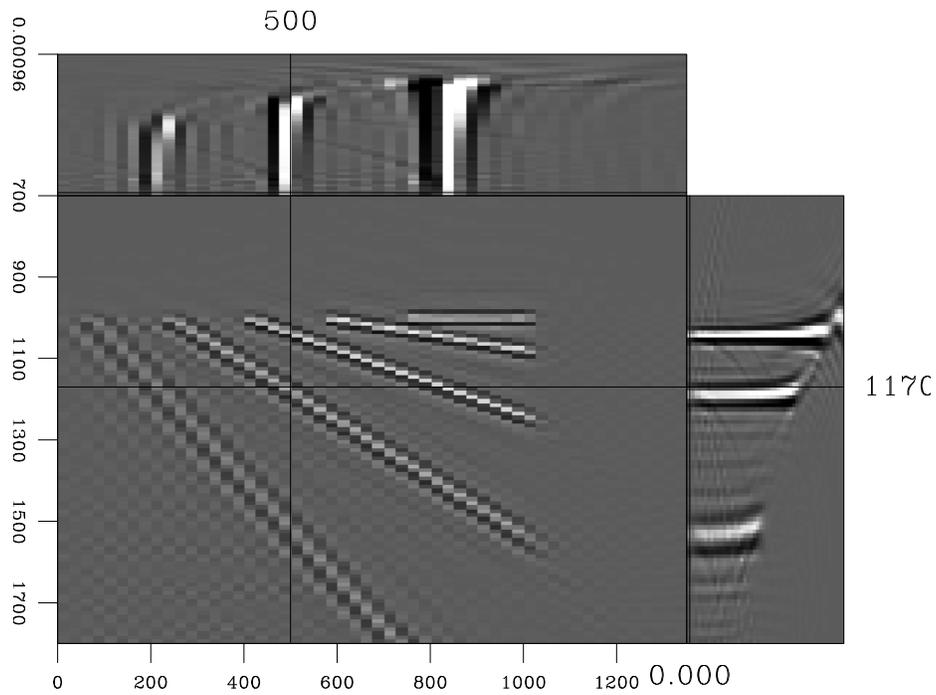


Figure 9: 3-D cube extracted from the 4-D image cube migrated with CAM, at location $\text{cmp-y}=500\text{m}$. Central panel (cmp-x / depth) shows all 5 dipping planes. Ray parameter domain CIGs show flat gathers except for the reflector dipping at 60° . `louis1-CA-mig-sect1` [CR]

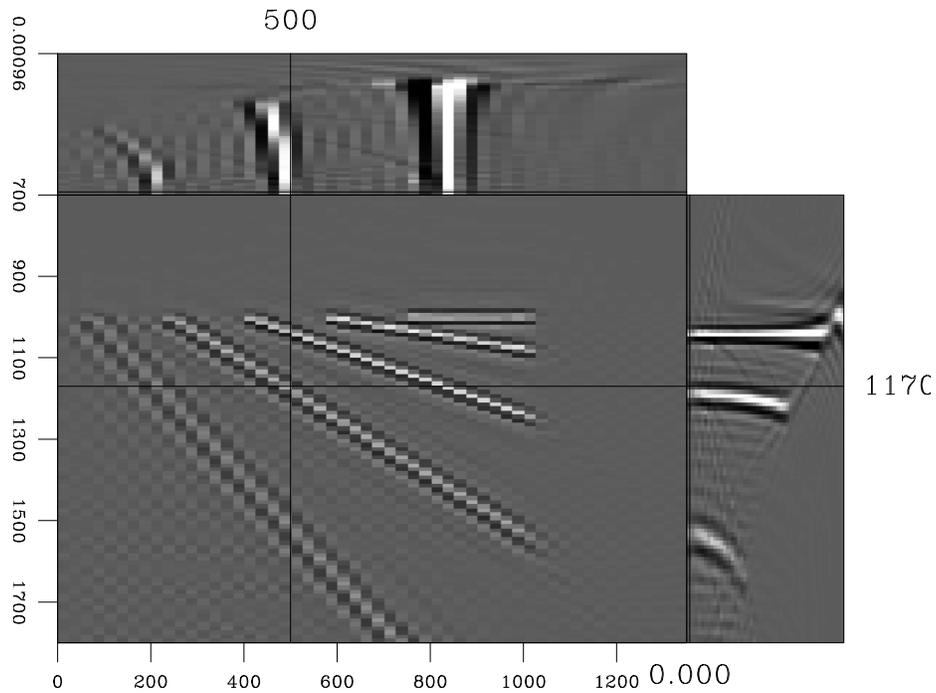


Figure 10: 3-D cube extracted from the 4-D image cube migrated with Offset-plane waves migration, at location $\text{cmp-y}=500\text{m}$. Ray parameter domain CIGs show perturbed gathers for reflectors dipping at 45° and 60° . `louis1-OPW-mig-sect1` [CR]

CIGs are flat for the first reflectors (dips $0 - 30^\circ$) in both images. For OPW migration, non-flat gathers start at dip 45° , whereas only a dip of 60° causes trouble to CAM.

In order to have a reference for comparison, we migrated the 4-D common-azimuth data with 5-D phase-shift migration, after zero-padding along a fictitious cross-line offset axis (Figure 8). Figure 11 shows flat gathers even for the most strongly dipping reflectors.

For using the full 5-D phase-shift operator, we added a fictitious cross-line offset axis by zero-padding common-azimuth data. The “arbitrary” parameters n_{hy} and d_{hy} are chosen in order to avoid wraparound problems in Fourier Transforms (n_{hy} large enough) and to have the exact value of k_{hy} included in our k_{hy} range. In practice, we used $n_{hy} = 24$ and $d_{hy} = 50m$.

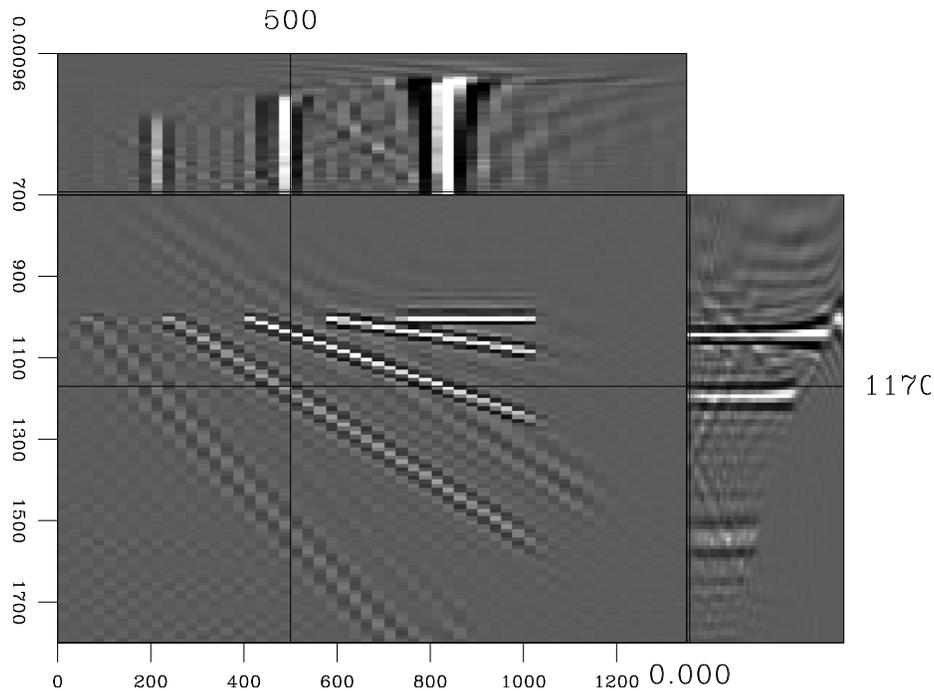


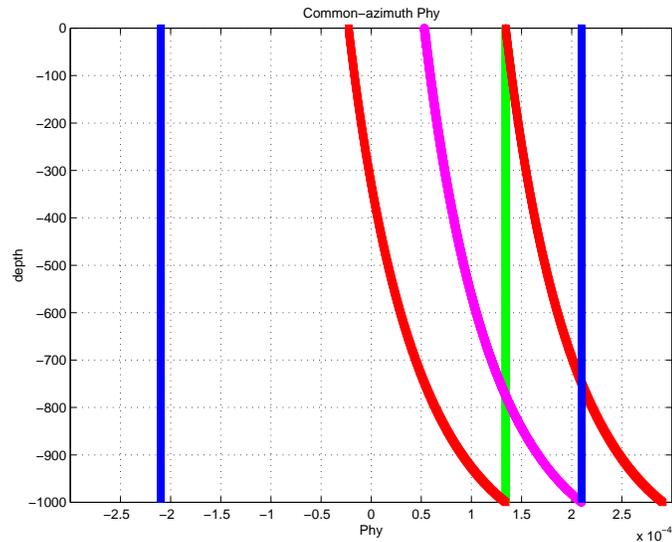
Figure 11: 3-D cube extracted from the 4-D image cube migrated with a 5-D phase-shift operator, at location $\text{cmp-y}=500m$. Ray parameter domain CIGs show completely flat gathers even at steep dips, as expected. [louis1-PS-mig-sect1](#) [CR]

The road to Narrow-Azimuth Migration (NAM)

Vaillant and Biondi (1999) reviewed common-azimuth migration theory and examined how to extend the method to a “narrow” range of azimuths. The previous discussion illustrates opportunities for obtaining the accuracy of the full 5-D phase-shift operator at a lower cost. Effectively, most of the contributions to the final image are concentrated in a cross-line offset wavenumber k_{hy} centered around CAM stationary path \hat{k}_{hy} . Summing all contributions coherently in such a narrow range (see Figure 12) can reduce the cost of applying the full 5-D phase-shift operator by a factor of about 5, with potentially the same accuracy at all dips.

Figure 12: Same reflector as in Figure 6. The dashed curve represents the stationary path \hat{k}_{hy} , with the estimated range needed for narrow-azimuth migration on the sides (dotted curves). The solid grey line is the exact value of k_{hy} . Black solid lines represent the minimum range needed for the full 5-D phase-shift operator.

`louis1-phy_nam` [ER]



CONCLUSIONS AND UPCOMING WORK

The stationary-phase approximation used to derive the CAM algorithm yields constraints in the downward-continuation process that limit accuracy in imaging, even in 1-D $v(z)$ media. However, those errors only become apparent for really steep reflectors (about 60°) with an important cross-line component. CAM is also a more reliable approximation of the 5-D phase-shift operator than the offset plane-wave algorithm. An even more robust approximation of the 5-D operator can be the narrow-azimuth migration algorithm, presently in progress. Eventually, 1-D media only allow us to estimate errors due to varying azimuth in downward-continuation. We will address issues related to multipathing and lateral velocity variations using the SEG/EAGE salt dome dataset.

REFERENCES

Vaillant, L., and Biondi, B., 1999, Extending common-azimuth migration: SEP-100, 125–134.

APPENDIX

Here we derive the equation of circular rays in constant gradient velocity media. Along a circular ray, the curvilinear abscissa s is, with the previous notations:

$$ds = R d\theta \quad (\text{A-1})$$

The equations for the ray trajectory passing through points $S(\rho_s, z_s, \theta_s)$ and $P(\rho, z, \theta)$ are

$$\begin{aligned}\rho - \rho_s &= \int_{\rho_s}^{\rho} d\rho' = \int_{\theta_s}^{\theta} \frac{d\rho'}{ds} \frac{ds}{d\theta'} d\theta' = \int_{\theta_s}^{\theta} R \sin \theta' d\theta' = \frac{1}{\gamma p_\rho} \int_{\theta_s}^{\theta} \sin \theta' d\theta' \\ &= \frac{1}{\gamma p_\rho} (\cos \theta_s - \cos \theta)\end{aligned}\quad (\text{A-2})$$

$$\begin{aligned}z - z_s &= \int_{z_s}^z dz' = \int_{\theta_s}^{\theta} \frac{dz'}{ds} \frac{ds}{d\theta'} d\theta' = \int_{\theta_s}^{\theta} R \cos \theta' d\theta' = \frac{1}{\gamma p_\rho} \int_{\theta_s}^{\theta} \cos \theta' d\theta' \\ &= \frac{1}{\gamma p_\rho} (\sin \theta - \sin \theta_s)\end{aligned}\quad (\text{A-3})$$

By writing the identity

$$\cos^2 \theta + \sin^2 \theta = [\gamma p_\rho (\rho_s - \rho) + \cos \theta_s]^2 + [-\gamma p_\rho (z_s - z) + \sin \theta_s]^2 = 1 \quad (\text{A-4})$$

at every point along the ray, we obtain the equation of the desired circle:

$$\left[\rho - \rho_s - \frac{\cos \theta_s}{\gamma p_\rho} \right]^2 + \left[z - z_s + \frac{v(z_s)}{\gamma} \right]^2 = \left[\frac{1}{\gamma p_\rho} \right]^2 \quad (\text{A-5})$$

Sub-salt imaging of a 2-D elastic synthetic model, using prestack, split-step, wave equation migration

*Douglas Gratwick*¹

ABSTRACT

This paper explores methods to image complex structures under a salt body, using a 2-D elastic synthetic model. The modeling algorithm handles complex features such as energy from multiple reflections and mode conversion. I show that a complex wave equation depth migration algorithm is needed to position the salt boundaries and reflectors from sediments below the salt properly. Also, using simple ray-tracing diagrams, I show that energy from converted waves is useful in imaging steeply dipping reflectors below the salt. Though these techniques improve overall image quality, parts of the image under the salt are still left unresolved, and I outline possible techniques for imaging these reflectors.

INTRODUCTION

The use of synthetic models is invaluable in our attempt to better process reflection seismic data (Audebert et al., 1994). The obvious benefit of these models is, of course, that they immediately provide the answer to the underlying question: What is under the surface where the survey is being done? Since the answer is known, individual problems can be isolated. Thus, if the project focus is on testing prestack migration and a synthetic model is used, the known velocity function is exactly correct. On the other hand, if the prestack migration algorithm works, then velocity functions can be modified to find how bad the input velocity function can be before the image is severely distorted (Etgen, 1990; van Trier, 1990; Biondi, 1990).

In this paper, only the prestack migration algorithm will be analyzed. Specifically, problems encountered with the imaging, and some techniques for solving those problems, will be covered. The model used for this project is complex, and thus provides for many interesting imaging problems. It will be shown that for such a complex setting, simple zero-offset migration techniques are not effective, and that more complex migration techniques which account for lateral velocity variations and depth conversion are necessary for a good image. In addition, the paper will demonstrate that even this sophisticated migration algorithm cannot take care of all imaging problems, and that other techniques need to be used in conjunction with the prestack migration algorithm. Specifically, a technique of changing the input velocity function will be implemented. By changing the input velocity of the salt from P-wave to S-wave,

¹email: doug@sep.stanford.edu

reflections from this wavefield are stacked in the image.

2-D SYNTHETIC MODEL

The data used in this project was provided by BP-Amoco, and is a 2-D, fully elastic synthetic model which has a number of complex features that can inhibit good imaging when using conventional methodology. In Figure 1, the P-wave velocity function is shown. The most dominant feature is the large salt body, which stretches for almost 50,000 feet, starting at 15,000 feet. It thickens to a maximum depth of over 3,000 feet. The salt is faulted on the base, and underlying the salt are a series of faulted blocks. Two layers can be seen in the fault blocks, with the layers being anomalously low velocity layers to the right, and slightly high velocity layers to the left. The high velocity layers become more contrasted towards the left of the model. The low velocity layers remain highly contrasted to the right edge of the model.

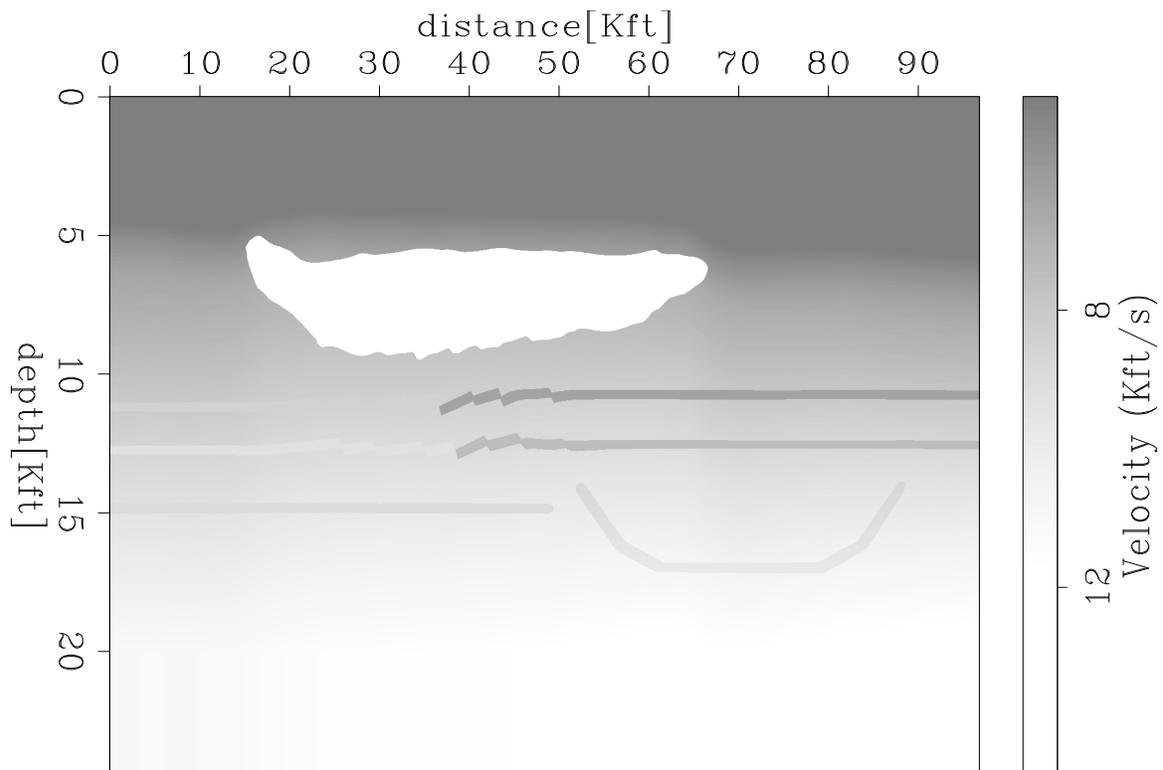


Figure 1: P-Wave Velocity Model `doug1-v_p` [ER]

Also shown is a synclinal structure (represents a paleochannel) below the two layers mentioned above, as well as a flat lens of sediment which starts on the left edge and stops abruptly underneath the salt. Both the syncline and lens are relatively low velocity layers. The S-wave velocity model seen in Figure 2 shows the same features, except that all the layers are consistently high velocity layers, and the layers in the faulted blocks remain at a constant velocity instead of abruptly changing as in the P-wave model.

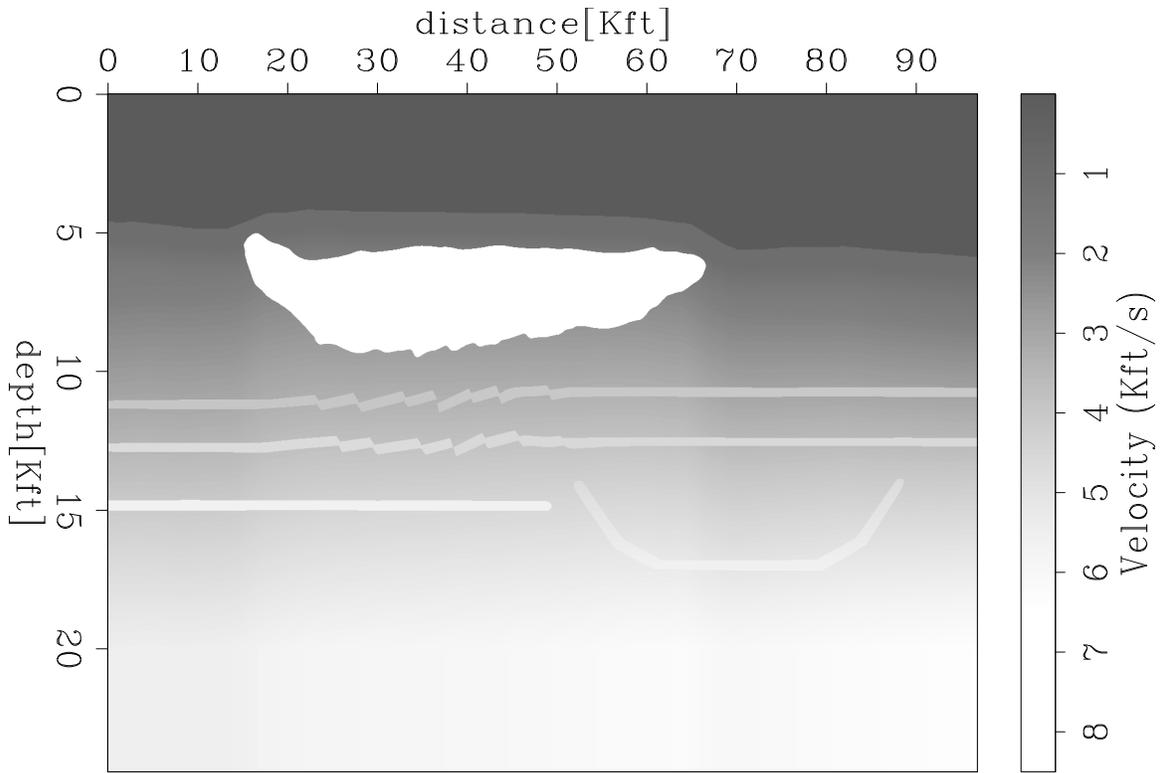


Figure 2: S-Wave Velocity Model `doug1-v_s` [ER]

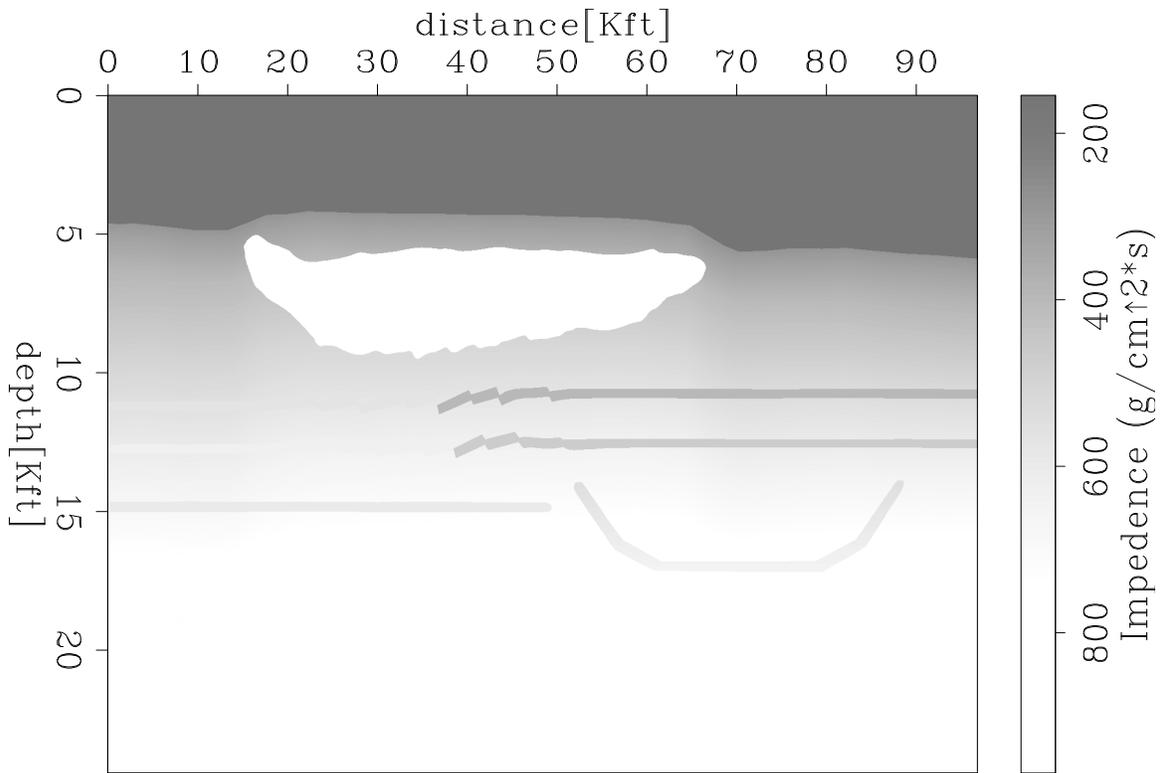


Figure 3: P-Wave Impedance Model `doug1-ip_bar` [ER]

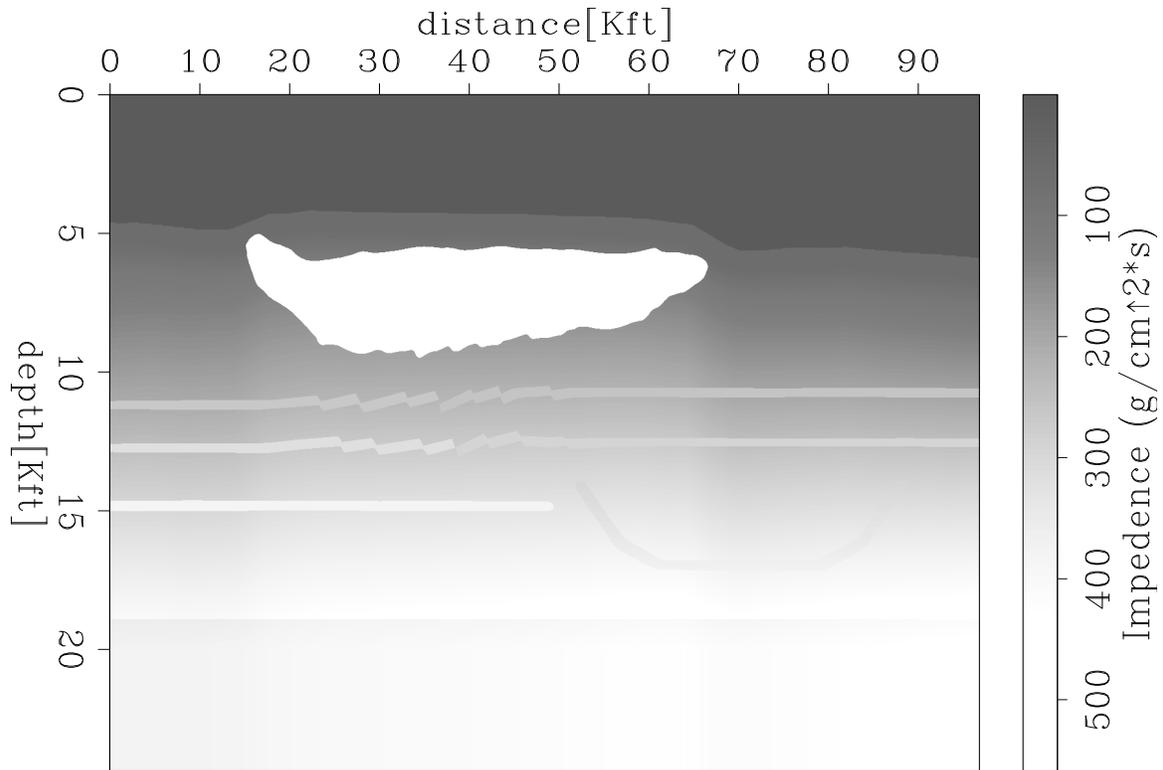


Figure 4: S-Wave Impedance Model `doug1-is_bar` [ER]

Impedances for the two models are shown in Figures 3 and 4. In the P-wave impedance model, large contrasts are seen for all of the features noted above, with the exception of the two slightly higher velocity layers seen at the left of the model, which end under the salt. In the S-wave impedance model, the impedance contrast remains constant for the two layers. A formula for reflectivity that is based on impedances is

$$R = \frac{Z_2 - Z_1}{Z_1 + Z_2} \quad (1)$$

where R is the reflectivity, and Z_1 and Z_2 are the impedances of the two layers.

From the formula, we can see that the magnitude of the reflection is independent of the position of the layers relative to each other. Rather, it is the sign which changes. So for high impedance over low, the reflection is in the same phase, while for low impedance over high, there is a 180° phase shift (Scheriff and Geldart, 1995). For the purposes of imaging, the strength of the reflection becomes important, simply because stronger reflection coefficients mean that the layers will have higher amplitudes in the image. Referring back to Figures 3 and 4, we can see that P-wave energy should be strong from the top and bottom of the salt, as well as from all the other layers, with the exception of the two high velocity layers. However, these layers show a higher S-wave impedance contrast, and thus S-wave energy should be reflected relatively well.

MIGRATION ALGORITHM

The migration algorithm used for this project uses a variation of the split-step Fourier domain migration method, developed by Stoffa (Stoffa et al., 1990). Simple Fourier domain migration based on the single square root operator (SSR) is also known as phase-shift migration (Gazdag, 1978). In phase-shift migration, a wavefield at the surface is downward-continued, assuming the wavefield is generated by the subsurface reflectors exploding (Claerbout and Black, 1997). The single square root operator marches the wavefield down, by shifting in the $\omega - k_x$ domain:

$$e^{ik_z \Delta z} = e^{-i \frac{2\omega}{v} \sqrt{1 - \frac{v^2 k_x^2}{4\omega^2}} \Delta z} \quad (2)$$

The imaging principal sums over all ω , giving the wavefield at $t=0$, which is the migrated image (Yilmaz, 1987). The split-step methods developed by Stoffa use an approximation of the SSR, where v is replaced by an average velocity of a given depth step, v_{ref} . A new term is added to correct for the difference of the average velocity and the actual medium velocity v_m . The equation looks like this variation of equation (2) (k_z is shown):

$$k_z = \sqrt{\frac{4\omega^2}{v_{ref}^2} - k_x^2} + \left(\frac{2\omega}{v_m} - \frac{2\omega}{v_{ref}} \right) \quad (3)$$

In equation (3), the first term is calculated in the $\omega - k_x$ domain, the second in the $\omega - x$ domain. So at every depth step, a 2-D FFT transforms the wavefield into $\omega - k_x$ space. The wavefield is downward-propagated using the phase-shift operator and $v=v_{ref}$. A 1-D IFFT transforms into $\omega - x$ space, where the correction term is applied. Once again, summing over all frequency gives the image. This method is exact if $v_{ref} = v_m$, or if layers are flat. If lateral velocity variations are significant or units are dipping steeply, a better approximation of the SSR equation is needed. To handle these conditions, we use the double square root equation (DSR). The DSR equation is a higher order approximation of the SSR equation. The DSR equation as displayed in IEI (Claerbout, 1984) is

$$\frac{dU}{dz} = -\frac{i\omega}{v} [\sqrt{1 - (Y + H)} + \sqrt{1 - (Y - H)}] U \quad (4)$$

where,

$$Y = \frac{v^2 K_y^2}{4\omega^2} \quad (5)$$

and

$$H = \frac{v^2 K_h^2}{4\omega^2} \quad (6)$$

The DSR equation inherently includes both migration and NMO. Therefore, with input CMPs and a velocity function, a good image is created. The actual migration program used in

this project is one developed by Biondo Biondi that uses a variation of the split-step method with the DSR equation (Biondi, 1998). The DSR equation can better handle lateral velocity variations as well as steeper dips (as compared to the SSR equation). The use of the split-step method with the DSR makes the algorithm more effective. In this way, multiple reference velocities can be used, and sharp lateral velocity variations, such as salt in contact with sediment, can be handled. For this project, three reference velocities were used. Biondi's program also has an input salt velocity, so that the reference velocities in the sediment will not be anomalously high if the salt is at the same depth step. So for each depth step, three wavefields were generated, then an interpolation in the space domain was used as needed.

ZERO-OFFSET MIGRATION

The main problem with the complex Fourier migration algorithm described above is that it is computationally expensive. However, by looking at the zero-offset section in Figure 5, we see that simple poststack migration will not be satisfactory. Since the salt body is present, a depth migration will help to restore the salt to its correct thickness. Also, with a more complex velocity function, there is a better chance of stacking out a majority of the multiple energy, which is very dominant in the CMP stack. In poststack migration, the object is to collapse diffractions, steepen dips, and uncross bow ties.

Using a poststack phase-shift operator, we can put the dips of the synclinal structure into the correct position, and the bow ties can be uncrossed. However, that is only a very small part of the image. Moreover, the times of those events are unlikely to be correct, especially under the salt. After migration it is still impossible to see under the salt, where the fault blocks are. Therefore, a complex prestack algorithm, though computationally expensive, becomes necessary in this type of complex geologic setting.

PRESTACK MIGRATION WITH P-WAVE VELOCITY MODEL

Initially, the most logical velocity function to input for prestack migration is that of the P-wave model. The energy which ends up in the gathers is usually dominated by P-waves, mainly because in this case, the source is P-wave and the receivers are in the water. So in order to get S-waves, mode conversions must occur. In fact a number of conversions must occur to get P-Waves which convert to S and then convert back to P. Conversion at interfaces reduces the energy, so after a number of conversions, not much energy is left (Scheriff and Geldart, 1995).

The section produced from the input CMPs and the P-wave velocity function is seen in Figure 6. This is an obvious improvement over the stacked section in Figure 5. The salt boundary has been imaged very well, including the faults at the base of the salt. Under the salt, the fault blocks to the right are imaged very well, with the two layers seen in the impedance plots very apparent in the image. However, as expected, the blocks and layers to the left are not very well imaged, because the impedance contrasts are such that not much energy is reflected.

We can also see that the part of the synclinal structure under the salt is not visible. The

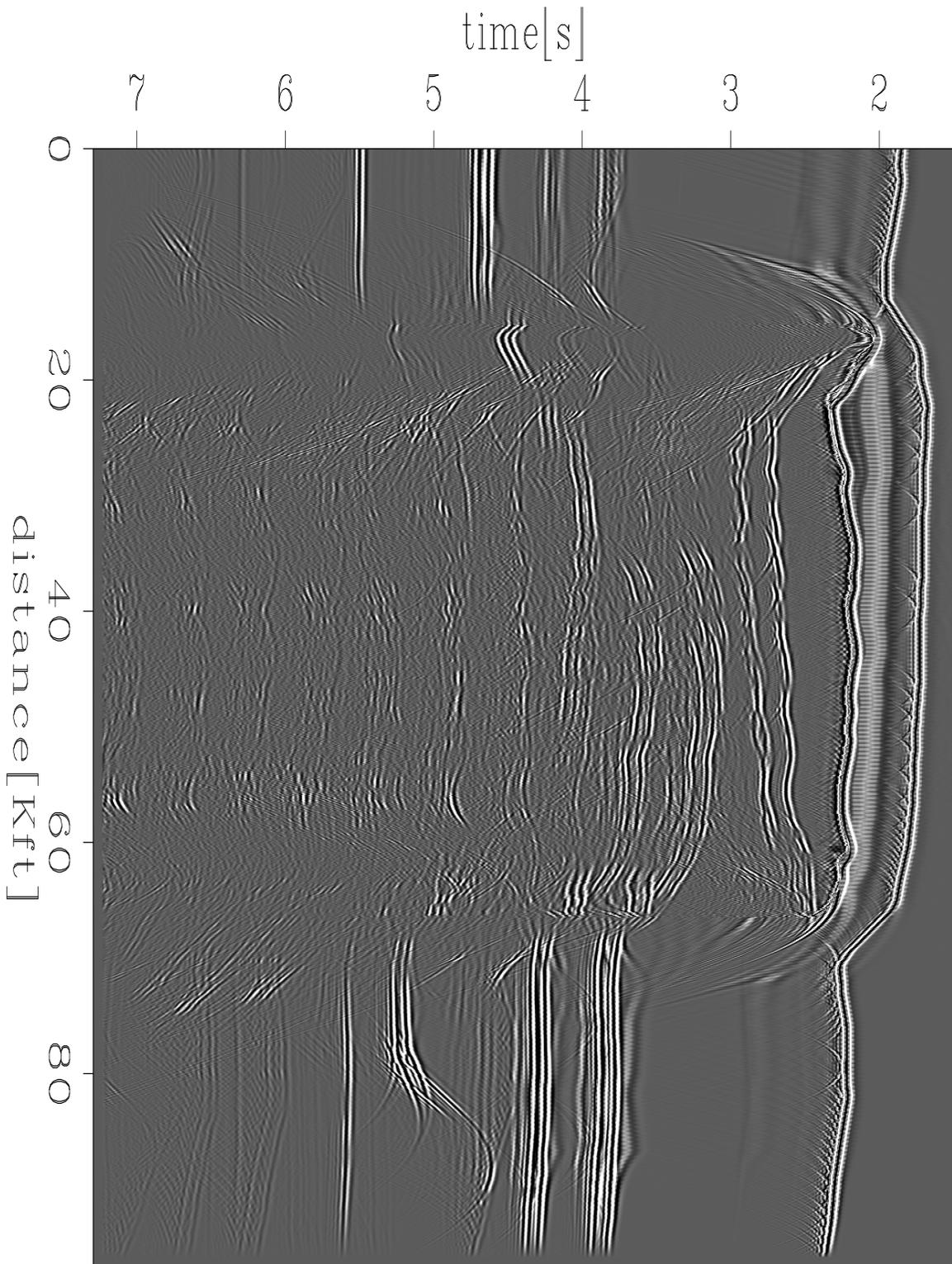


Figure 5: Zero-Offset Section `doug1-stack` [CR]

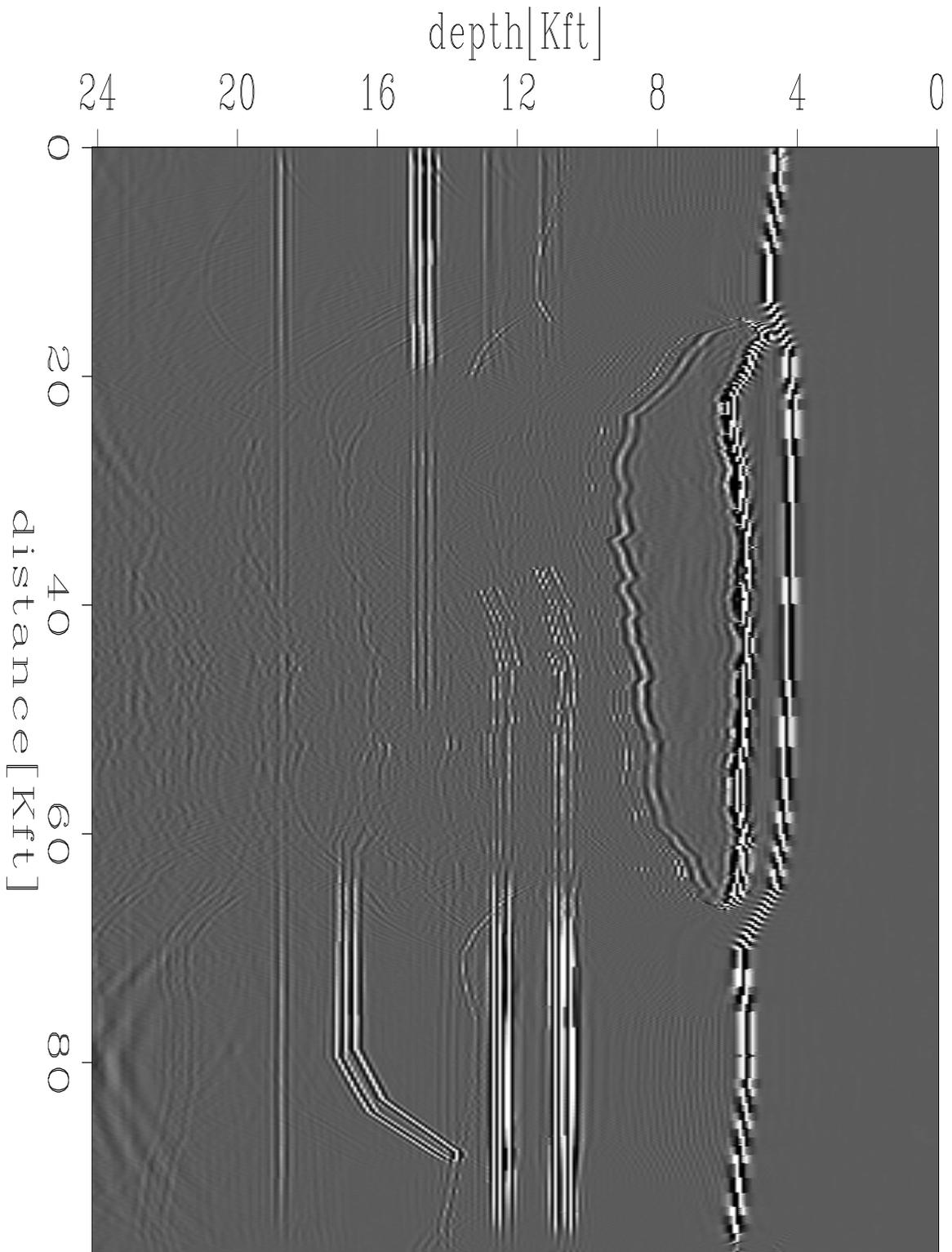


Figure 6: Prestack Migrated Section Using P-Wave Velocity Function `doug1-prestack.p` [CR]

dips and positions of the right flank seem good, but under the salt, the image is lost. This is likely due to a problem with the raypaths being distorted as the wavefield is going through the salt. This causes less energy to get to the receivers, and thus that part of the image does not show up (Muerdter et al., 1996). The same scattering effects are seen in the lens, which has a shadow zone under the salt boundary where the wavefield energy does not reach the receivers. Another feature is the trace of the salt bottom which appears just a little bit down from the bottom of the salt. This is actually the S-wave energy from the travel of converted waves in the salt which stacks at that velocity. With the exception of these shadow zone problems, and some of the S-wave energy, the image is very good when compared to the actual model.

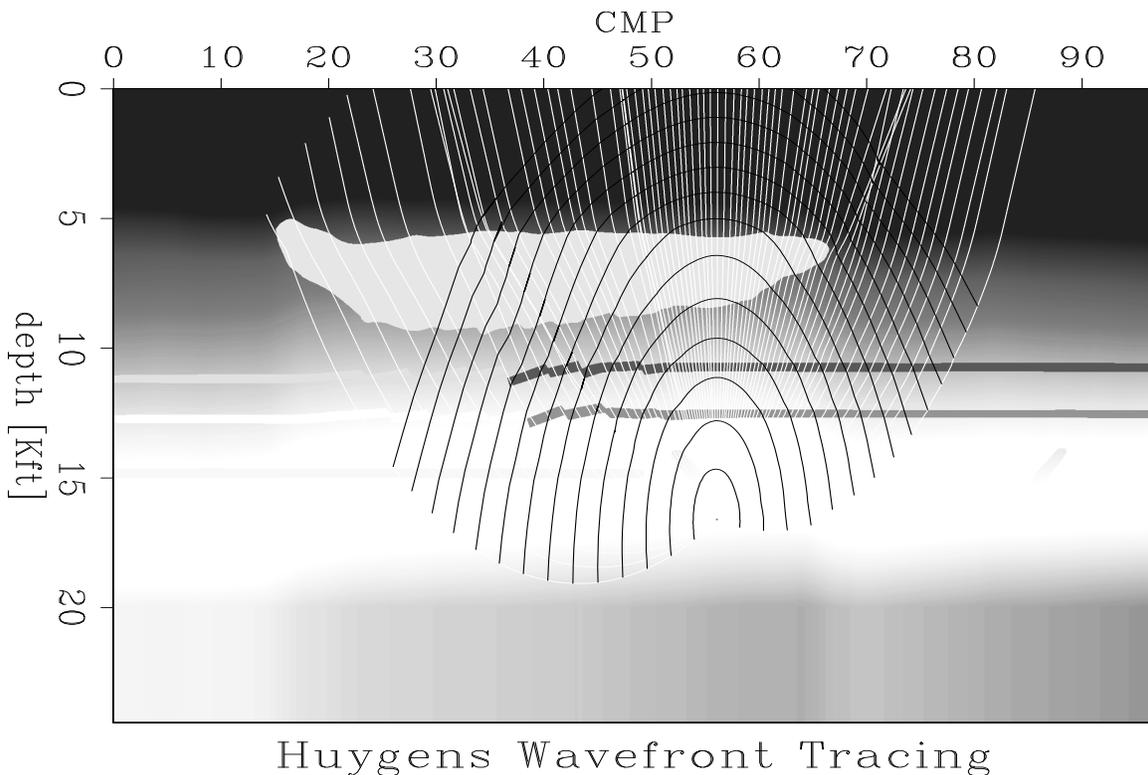


Figure 7: Wavefield Trace Of P-Wave Model With S-Wave Salt Velocity doug1-vwp.saltlow
[ER]

PRESTACK MIGRATION WITH MODE-CONVERTED VELOCITY MODEL

Note that the scattering effects of the salt cause much of the P-wave energy to be lost when we are dealing with structures under the edge of the salt body. Moreover, when we consider the steep dip of the synclinal structure, it is no surprise that very little P-wave energy gets to the surface from these areas. The next step is then to change the velocity model to stack different arrivals that have a better chance to make it to the surface (Kessinger, 1996). These arrivals are the ones that are from the wavefield that converts to S-wave through the salt. The interface of the sediment and salt is such that much of converted energy can be produced from an incident

P-wave. Figure 7 shows a ray trace of a wavefield generated on the left flank of the synclinal structure. Note that the wavefield does not scatter through the salt body, which has the S-wave salt velocity, while the sediment has the P-wave velocity.

In the migration that used this new velocity model, we can see that enough of this converted wave energy gets to the receivers that the left flank of the synclinal structure is imaged. Although its amplitude is not very strong, it is enough that an interpreter would recognize the feature. Note also that the trace below the salt bottom that was seen in the previous migrated image has moved to its correct subsurface position (Figure 8). Still not imaged well, however, are the leftmost fault blocks under the salt body, as well as the shadow zone of the lens described in the previous section.

CONCLUSIONS/FUTURE CONSIDERATIONS

It is clear that the migration algorithm described and implemented in this paper does a good job of imaging in the presence of complex geology. The result from the prestack migration vs. the poststack migration shows that in this case, the use of the prestack migration algorithm was not only beneficial, it was necessary. However, the limitations of the use of this procedure by itself are shown in its failure to image certain parts of the model. Though simple tweaking of the velocity model can prove to help image different parts, others were not visible. The faulted blocks with the relatively high S-wave impedance contrasts are symbolic of a real life possibility where the layers are brine-bearing sediments. In this case, these layers are important things to image, so that they are avoided while drilling. To image these parts, we would need some special processing to take place. One method that could be used is multiple attenuation. If the multiples from the water bottom and salt were adequately attenuated, the gain could be turned up and the low amplitude signals might be more easily seen. Another method is to mute the energy in the data so that only the target reflections (deeper units of interest) would be imaged.

Currently, we are considering the latter, and there is initial evidence that the S-wave energy from the left fault block layers is enough that it can be identified and separated from the rest of the energy on the CMP gathers. If this can be done for all the CMPs which correspond to that part of the image, there is a good possibility that we could image these blocks.

REFERENCES

- Audebert, F., Biondi, B., Lumley, D., Nichols, D., Rekdal, T., and Urdaneta, H., 1994, Marmousi traveltimes computation and imaging comparisons: *SEP-80*, 47–66.
- Biondi, B., 1990, Velocity analysis using beam stacks: Ph.D. thesis, Stanford University.
- Biondi, B., 1998, 3-D seismic imaging: *SEP-98*, 1–204.
- Claerbout, J. F., and Black, J. L., 1997, Basic Earth Imaging: Class notes, <http://sepwww.stanford.edu/sep/prof/index.html>.

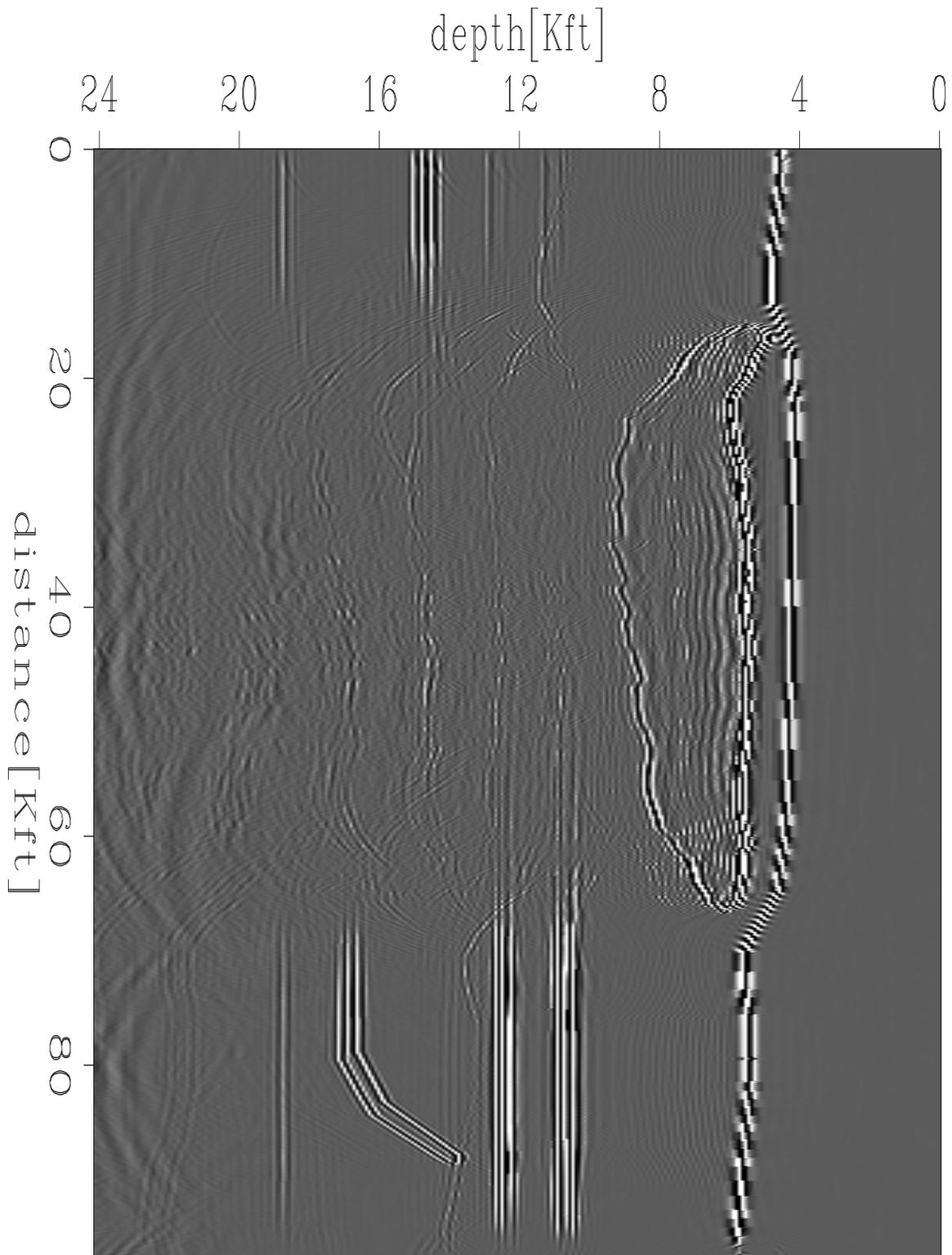


Figure 8: Prestack Migrated Section Using P-Wave Velocity Model with Lowered Salt Velocity `doug1-prestack.p.saltlow` [CR]

- Claerbout, J. F., 1984, Imaging the Earth's Interior: SEP-40.
- Etgen, J., 1990, Residual prestack migration and interval velocity estimation: Ph.D. thesis, Stanford University.
- Gazdag, J., 1978, Wave-equation migration by phase shift: *Geophysics*, **43**, 1342–1351.
- Kessinger, W. e. a., 1996, Subsalt imaging using mode converted energy and acoustic depth migration: 66th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts with Authors, v.1 566–569.
- Muerdter, D. R., Lindsay, R. O., and Ratcliff, D. W., 1996, Imaging under the edges of salt sheets: a raytracing study: 66th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 578–580.
- Scheriff, R., and Geldart, L., 1995, *Exploration seismology*: Cambridge.
- Stoffa, P., Fokkema, J., de Luna Freire, R. M., and Kessinger, W. P., 1990, Split-step fourier migration: *Geophysics*, **55 no.4**, 410–421.
- van Trier, J., 1990, Tomographic determination of structural velocities from depth migrated seismic data: Ph.D. thesis, Stanford University.
- Yilmaz, O., 1987, *Seismic data processing*: Soc. Expl. Geophys.

Prestack multiple attenuation using the hyperbolic Radon transform

A comparison of inversion schemes

*Antoine Guitton*¹

ABSTRACT

I apply the iterative hyperbolic Radon transform to CMP gathers to create a velocity panel where multiples and primaries are separable. The velocity panel is created using three different inversion schemes: (1) l^2 norm inversion, (2) l^1 norm inversion and (3) l^1 norm with l^1 regularization inversion. The third technique is particularly efficient at separating primaries and multiples in the prestack domain. A comparison of the three techniques shows that some noticeable differences appear in the prestack domain after multiple attenuation and that no discrepancies emerge on the stacked sections. These conclusions are linked to convergence properties of each method, and also linked to the “quality” of the data.

INTRODUCTION

The last decade has seen an exponential growth in the use of 3-D seismic imaging. Contemporaneous with this development, imaging techniques have become more complex in the effort to account for multi-pathing in complex media and to produce “true amplitude” migrated pictures of the subsurface. Since multiples are not accounted for in the physical model that leads to these migration methods, they can severely affect the final migration result producing erroneous interfaces or amplitude artifacts; consequently, the multiples have to be removed from the data. As pointed out by Weglein (1999), the multiple attenuation techniques may be divided into two families: (1) filtering methods which exploit the periodicity and the separability (move-out discrepancies) of the multiples and (2) the wavefield prediction/subtraction methods, where the multiples are first predicted (Verschur et al., 1992; Weglein et al., 1997) and then subtracted (Spitz, 1999; Doicin and Spitz, 1991; Dragoset and MacKay, 1993; Clapp and Brown, 1999; Brown et al., 1999).

As oil companies lead exploration towards more complex geological structures (e.g., salt plays) and use 3-D surveys intensively, the attenuation of the multiples becomes more challenging. Spitz (2000, Personal communication) recently asserted that multiples are the number one problem in seismic processing. Traditionally, filtering techniques are the method of choice for multiple processing because of their robustness and cost. However, they have some

¹email: antoine@sep.stanford.edu

limitations when tackling multiples in complex media (predictive deconvolution) and in the preservation of primaries' amplitude (f - k filters). Wavefield methods overcome these limitations, therefore they are becoming more popular in the seismic industry. Nonetheless, they are often arduous to tune, generally slow, and very difficult to extend in 3-D for coverage reasons.

My approach

This paper will describe my production of a velocity domain where multiples and primaries are easily identifiable and separable by the use of the Hyperbolic Radon Transform (HRT) and inverse theory. The **Huber function** (Huber, 1973), or Huber norm, allows us to solve hybrid l^1 - l^2 inverse problems in an efficient fashion. I compare three different methods to obtain the velocity panel: (1) least-squares inversion, (2) l^1 inversion, and (3) l^1 inversion with l^1 regularization. These velocity panels are then used to perform the multiple suppression (Lumley et al., 1995).

In this paper I will first review the theory of the velocity transform operator. Next, I introduce the Huber norm and the inverse problem I intend to solve to produce the velocity field. Finally, I apply a multiple attenuation technique for different inverse problems to a complete 2-D data set (Mobil AVO data). I will show that the multiple reflections are favorably attenuated with no noticeable differences between the different inverse problems.

THEORY

Definitions of operators

The HRT maps the data (t, x) into a velocity space (τ, v) that clearly exhibits the moveout inherent in the data and, therefore, forms a convenient basis for velocity analysis. Thorson and Claerbout (1985) were the first to define the forward and adjoint operators of the HRT, formulating it as an inverse problem, where the velocity domain is the unknown space. In their approach, the forward operator \mathbf{H} stretches the model space (velocity domain) into the data space (CMP gathers) using a hyperbola superposition principle, whereas the adjoint operator \mathbf{H}^\dagger , the HRT, squeezes the data summing over hyperbolas (related to the velocity stack as defined by Taner and Koehler (1969)). The forward operation is

$$d(t, x) = \sum_{s=s_{min}}^{s_{max}} w_o m(\tau = \sqrt{t^2 - s^2 x^2}, s), \quad (1)$$

and the adjoint transformation becomes

$$m(\tau, s) = \sum_{x=x_{min}}^{x_{max}} w_o d(t = \sqrt{\tau^2 + s^2 x^2}, x), \quad (2)$$

where x is the offset, s the slowness, τ the two-times zero offset travel time, and w_o a weighting function that compensates to some extent for geometrical spreading and other effects (Claerbout and Black, 1997).

The inverse problem

Having defined the forward operator \mathbf{H} and its adjoint \mathbf{H}^\dagger , we can now pose the inverse problem. Inverse theory helps us to find a velocity panel which synthesizes a given CMP gather *via* the operator \mathbf{H} . In equations, given data \mathbf{d} (CMP gather), we want to solve for the model \mathbf{m} (velocity panel)

$$\mathbf{H}\mathbf{m} = \mathbf{d},$$

which is equivalent to the linear system

$$\mathbf{H}^\dagger\mathbf{H}\mathbf{m} = \mathbf{H}^\dagger\mathbf{d}.$$

This system is easy to solve if $\mathbf{H}^\dagger\mathbf{H} = \mathbf{I}$, i.e., if \mathbf{H} is unitary. Unfortunately, \mathbf{H} is far from an unitary operator. Sacchi and Ulrych (1995) give a couple of reasons for this behavior (see also Kabir and Marfurt (1999) for a more graphical interpretation of the artifacts):

1. The velocity (slowness) range is not wide enough.
2. The sampling in the velocity domain is too coarse.

Inverse problems are often used to handle the non-unitarity of operators. A prior advisable step in the design of an inverse problem is to attribute some properties to the model in terms of moments of corresponding distributions. A reasonable property of the model space would be **sparseness**, meaning that we want to cluster the components of the solution into a few large peaks (Thorson and Claerbout, 1985). The sparseness would help to distinguish primaries and multiples in the velocity space. Finally, we would like to design a solver that bears robustness to bad (inconsistent) data points. These bad data points leave large values in the residual and attract most of the solver's efforts (Fomel and Claerbout, 1995). Unfortunately, seismic data are generally very noisy, and the need for robust estimators is very pressing.

How to obtain a sparse model?

Some authors have proposed different solutions to address the sparseness of the model space. Thorson and Claerbout (1985) developed a stochastic inversion scheme that converges to a solution with minimum entropy. Sacchi and Ulrych (1995) apply a very similar method with more degrees of freedom to the choice of parameters. Nichols (1994) uses a regularization term with the l^1 norm. All these methods assign long-tailed density functions to the model parameters. Figure 1 shows an exponential (related to the l^1 norm) and a Gaussian distribution (related to the l^2 norm). The Gaussian distribution will tend to smooth the model space, spreading the energy, whereas the exponential distribution will tend to focus the energy on a few peaks, neglecting average values, and thus leading to a sparse model.

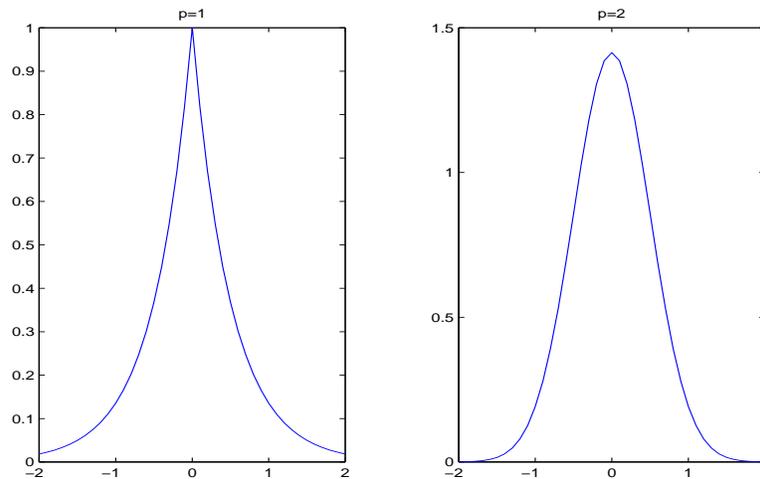


Figure 1: Exponential (left) and Gaussian (right) distribution with zero mean. The exponential distribution has the longer tail. [antoine1-distri](#) [CR]

How to design a robust solver

The need for a robust solver may be addressed using the l^1 norm for the data residual (Claerbout and Muir, 1973). Again, robust measures are related to the long-tailed density function in the same way that the mean square is related to the (short-tailed) Gaussian (Tarantola, 1987). The l^1 norm is then less sensitive to outliers and will give a more probable fitting of the data.

The requirements in the design of a robust inverse method that gives a sparse model for the velocity estimation problem leads to the minimization of the objective function

$$f(\mathbf{m}) = \|\mathbf{H}\mathbf{m} - \mathbf{d}\|_1 + \sigma \|\mathbf{m}\|_1, \quad (3)$$

where $\|\cdot\|_1$ is the l^1 norm. Since we wish to utilize the l^1 norm, the minimization of f is a cumbersome problem. The l^1 norm is not differentiable everywhere, which makes its use rather difficult. The next section presents some alternatives to the l^1 norm using hybrid l^1 - l^2 objective functions. These functions are differentiable and allow the use of iterative methods.

Hybrid l^1 - l^2 function

The hybrid l^1 - l^2 norm has been widely used in geophysics (Bube and Langan, 1997; Nichols, 1994; Fomel and Claerbout, 1995). It is generally solved using iteratively reweighted least-squares (IRLS) algorithms with an appropriate weighting matrix. These algorithms have proved efficient but are also acknowledged to be difficult to tune. As an alternative to using IRLS algorithms to compute the hybrid l^1 - l^2 norm, Claerbout (1996) introduced the **Huber norm** (Huber, 1973). This norm is a patching of the l^1 norm for high residuals and of the l^2

norm for small residuals:

$$M_\epsilon(r) = \begin{cases} \frac{r^2}{2\epsilon}, & 0 \leq |r| \leq \epsilon \\ |r| - \frac{\epsilon}{2}, & \epsilon < |r| \end{cases} \quad (4)$$

where r is the residual. We call $\sum_{i=1}^N M_\epsilon(r_i)$ the *Huber misfit function*, or the Huber function, for short (Figure 2). Note that the Huber function is smooth near zero residual and weights small residuals by the mean square. It is reasonable to suppose that the Huber function, while maintaining robustness against large residuals, is easier to minimize than l^1 . The parameter ϵ , which controls the limit between l^1 and l^2 , is called the Huber threshold.

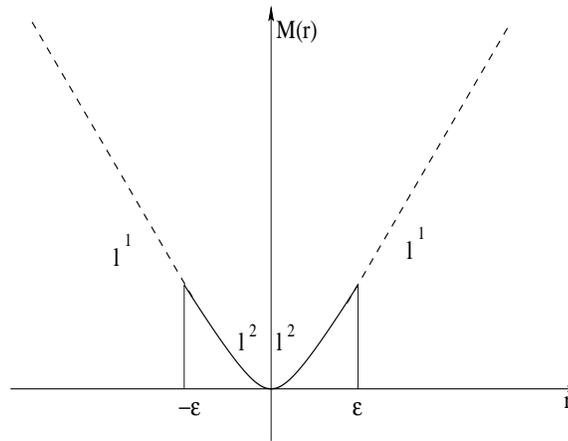


Figure 2: Error measure proposed by Huber (1973). The upper part above ϵ is the l^1 norm, while the lower part is the l^2 norm. antoine1-huber [NR]

The implementation of an inverse solver to minimize the Huber function is quite challenging and leads to innovative non-linear algorithms in geophysics (Guitton, 2000b). To summarize, I developed a quasi-Newton method with a line search. I implemented a More and Thunten Line Search (More and Thunten, 1994) algorithm, which ensures a sufficient decrease in the objective function f (equation 3) and obeys curvature conditions (the so-called *Wolfe conditions*, Kelley (1999)). The update of the Hessian is made using a Limited Memory BFGS method as proposed by Nocedal (1980) and Liu and Nocedal (1989). This method is guaranteed to converge to a minimum. This strategy has proved efficient in solving the Huber problem correctly (Guitton and Symes, 1999; Guitton, 2000a) and eliminates the restart parameter encountered in IRLS algorithms, which makes this *Huber solver* easier to use.

What will I do?

My goal in this paper is to compare three different inversion schemes for the multiples attenuation problem. They all aim to produce a velocity model where primaries are muted out and the predicted multiples are subtracted from the original data. I successively solve

1. $f(\mathbf{m}) = \|\mathbf{H}\mathbf{m} - \mathbf{d}\|_2$,
2. $f(\mathbf{m}) = \|\mathbf{H}\mathbf{m} - \mathbf{d}\|_1$,

$$3. f(\mathbf{m}) = |\mathbf{H}\mathbf{m} - \mathbf{d}|_1 + \sigma |\mathbf{m}|_1,$$

and compare the results. I call arbitrarily “ l^1 norm” any Huber function with a small threshold. Let us assume now that to the l^1 norm, for the data residual, corresponds a threshold

$$\epsilon = \frac{\max|\mathbf{d}|}{100}.$$

In addition, for the regularization term, let us say that to the l^1 norm corresponds a threshold

$$\epsilon = \frac{\max|\mathbf{d}|}{10000}.$$

ϵ is chosen smaller than before leading to a larger l^1 treatment of the model. I show later on that the convergence is greatly reduced by the addition of this regularization term.

MARINE DATA RESULTS

This section presents results obtained using the HRT on a the Mobil AVO data set (Lumley et al., 1995). These prestack data are heavily contaminated by free-surface and water-bottom multiples and thus constitute a challenging test-bed for true amplitude multiple attenuation techniques. The strategy for the multiple elimination is as follows:

1. Create a velocity panel with the iterative Hyperbolic Radon Transform using the Huber solver.
2. Define a corridor between multiples and primaries in the velocity space.
3. Mute out the primaries in the velocity space.
4. Model back the multiples in the data space.
5. Subtract the multiples field from the input data.
6. NMO and Stack.

This method is most suitable for preserving primaries energy. I compare l^2 , l^1 and l^1 with l^1 regularization for the inversion in step 1. I decide to parameterize the problem using velocities and not slownesses.

Computing aspects

The velocity-stack inversion has been fully automated, implying that a threshold is adaptively computed for each CMP gather independently. The input data are composed of 801 gathers

with 60 traces each. In my implementation, I parallelized the computing using OpenMP²(this concerns the operators only). On 10 processors of our SGI Power Challenge machine (notoriously slow), for 30 iterations, it took 30 hours to compute the 801 velocity panels. This process may be accelerated by faster machines.

Inversion results

Figure 3 shows the result of the inversion for one CMP. Since the Mobil AVO dataset does not include very complex structures with strong velocity contrasts, this panel illustrates what happens for all the gathers. The left panel shows the input data. The other panels display the reconstructed data using the different schemes. Note that the l^2 and l^1 inversions give similar results and that the l^1 regularization doesn't converge as well. Figure 4 highlights this difference between the different problems. The best convergence is achieved using least-squares and the worst is achieved with the l^1 regularization. In my implementation, however, the l^1 problem with or without regularization was solved using twice as many iterations as with l^2 . Figure 5 shows the differences between the input data and the remodeled data. It appears that the l^1 norm with l^1 regularization encounters some difficulties in fitting the far offset data. Note that the l^1 norm and the l^2 norm are both comparable. This is expected since the data are not strongly noisy.

Differences arise in favor of the l^1 regularization when we look at the model space (Figure 6), however. The l^1 and l^2 results are again very similar and the l^1 norm with l^1 regularization appears spikier. This result is consistent with the theory (see Theory section). The spiky result is then used to define the limit between primaries and multiples (black line in the right panel of Figure 6). A mask is defined accordingly and the primaries are muted out in the model space. The next step consists of remodeling the multiples back in the data space, applying the hyperbola superposition principle (operator **H**). Figure 7 shows the predicted multiples. Note that for the three inversion schemes, some primaries remain. This is particularly annoying to us in our attempt to produce true amplitude multiple-free gathers. Figure 8 displays the result of the multiple attenuation process. The three methods display similar results. Nonetheless, at far offset, the l^1 regularization shows more energetic events. This is consistent with Figure 5 where we showed that the l^1 regularization was unable to fit this part of the data.

NMO-Stacking process

In the last step, the data are NMO corrected and stacked. The stacking velocity comes from Lumley et al. (1995). Figure 9 shows the stack of the input data without multiple attenuation. Figures 10, 12, 14 show the stacks of the multiple-free data with the different methods. Figures 11, 13, 15 show the difference between the stacked section of the input data *with* multiples and the stacked section of the data *without* multiples for the three inversion schemes. We see that the multiple suppression has cleaned up deeper parts of the structure (below 2s.). Furthermore,

²OpenMP is a specification for a set of compiler directives, library routines, and environment variables that can be used to specify shared memory parallelism in Fortran and C/C++ programs (www.openmp.org).

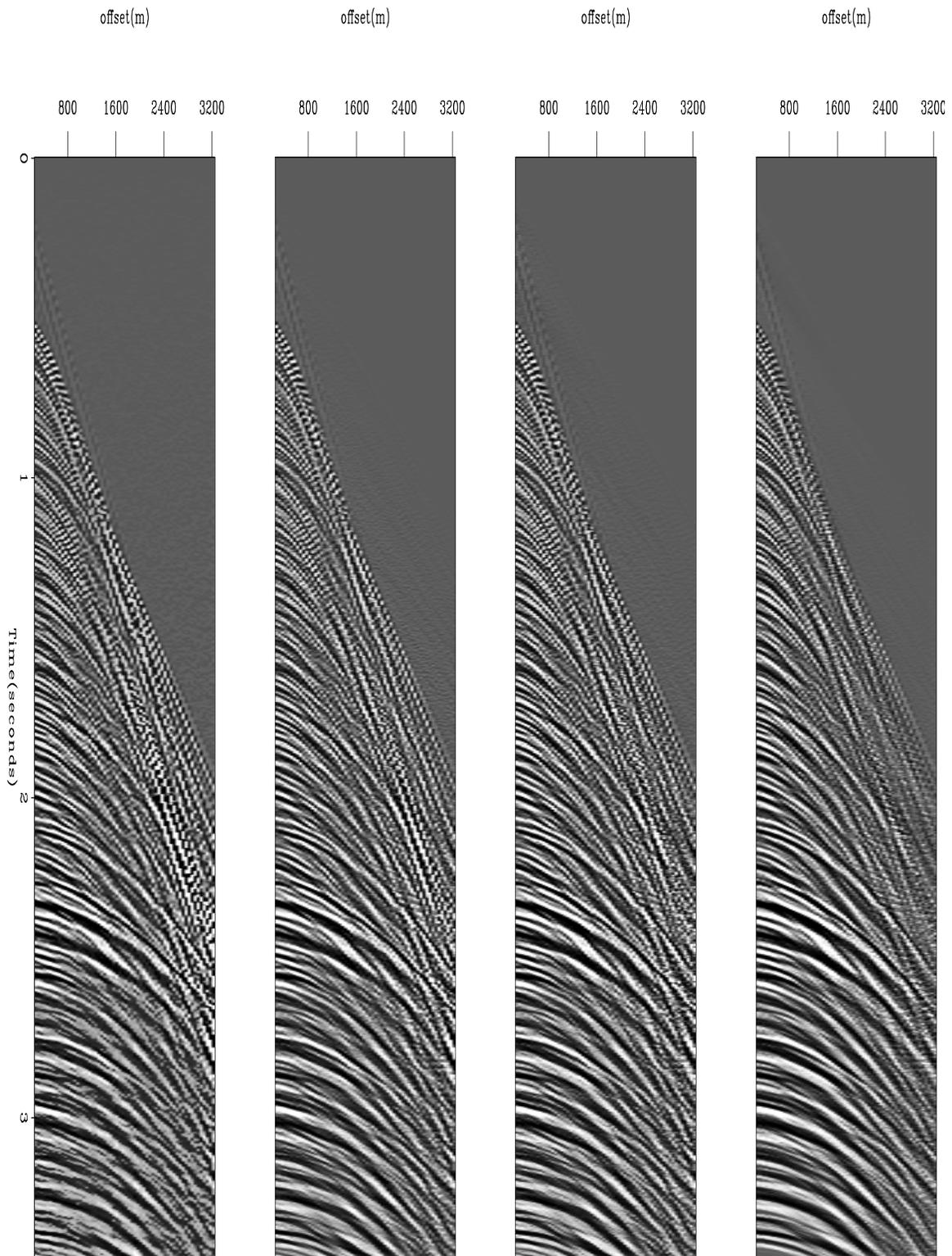


Figure 3: Left: input data. Middle-left: l^2 reconstructed data. Middle-right: l^1 reconstructed data. Right: l^1 with l^1 regularization reconstructed data. `antoine1-comp_dat` [CR]

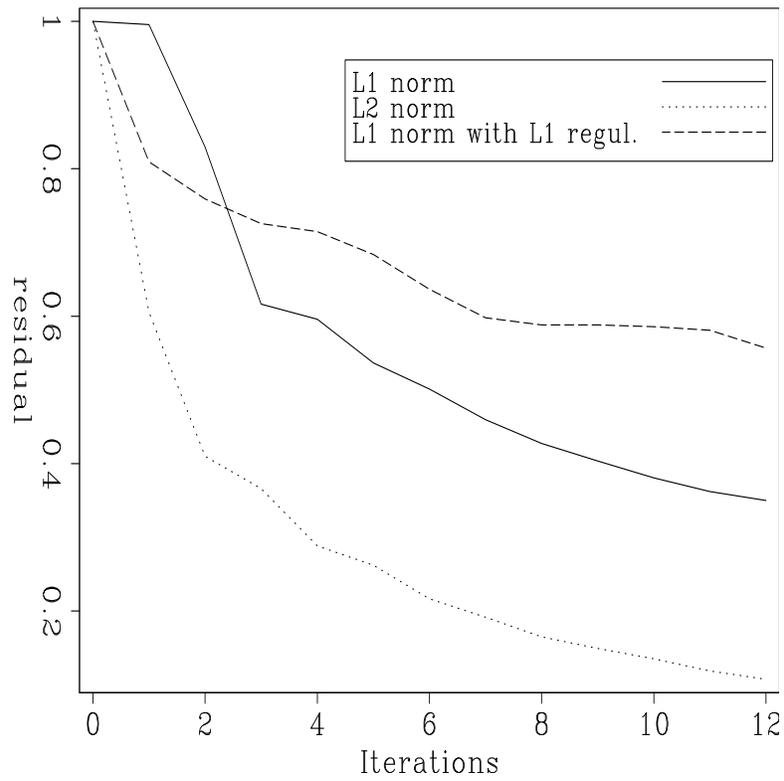


Figure 4: Comparison of the convergence for different inversion schemes for one CMP gather.
`antoinel-residual` [CR]

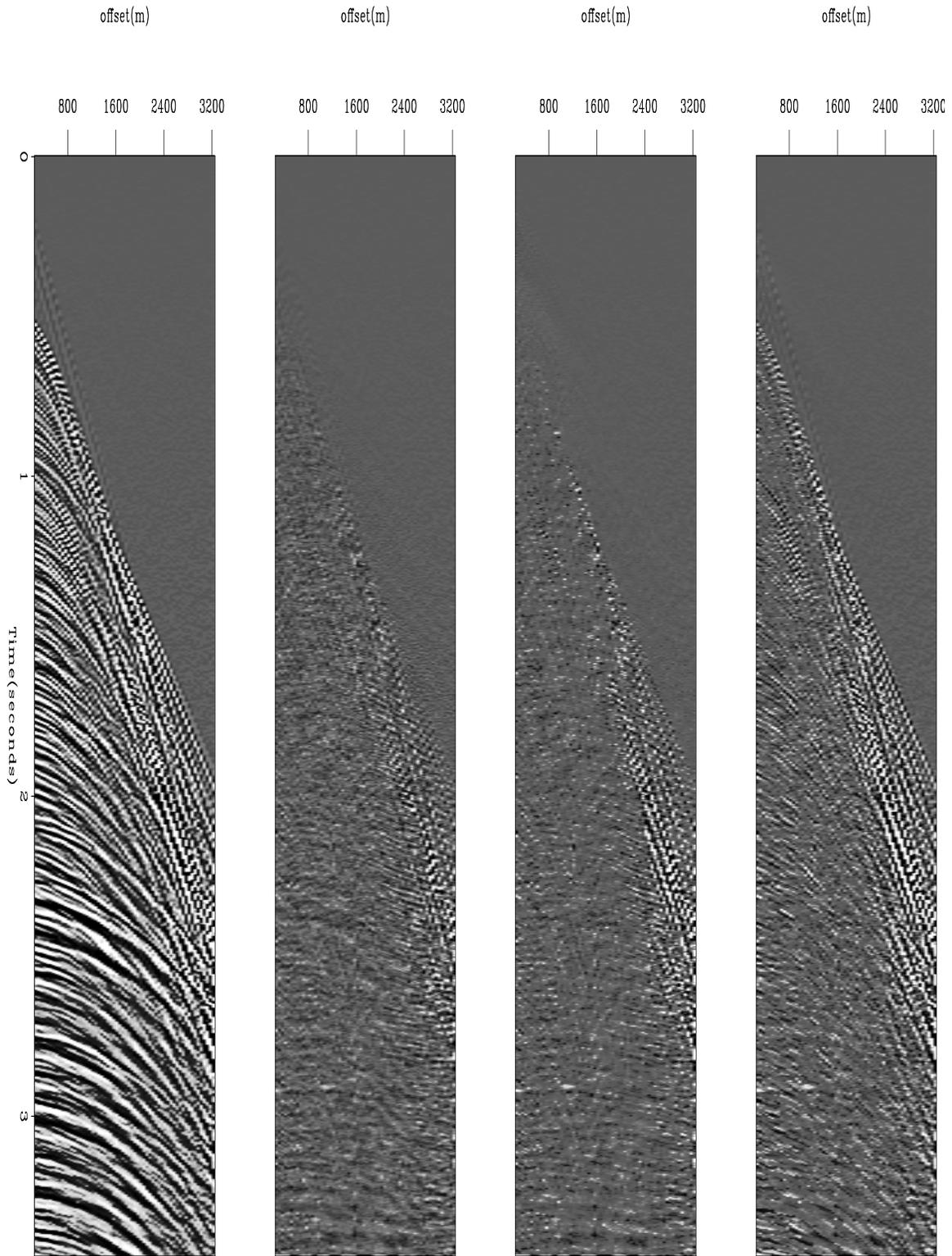


Figure 5: Left: input data. Middle-left: l^2 residual. Middle-right: l^1 residual. Right: l^1 with l^1 regularization residual. `antoine1-diff` [CR]

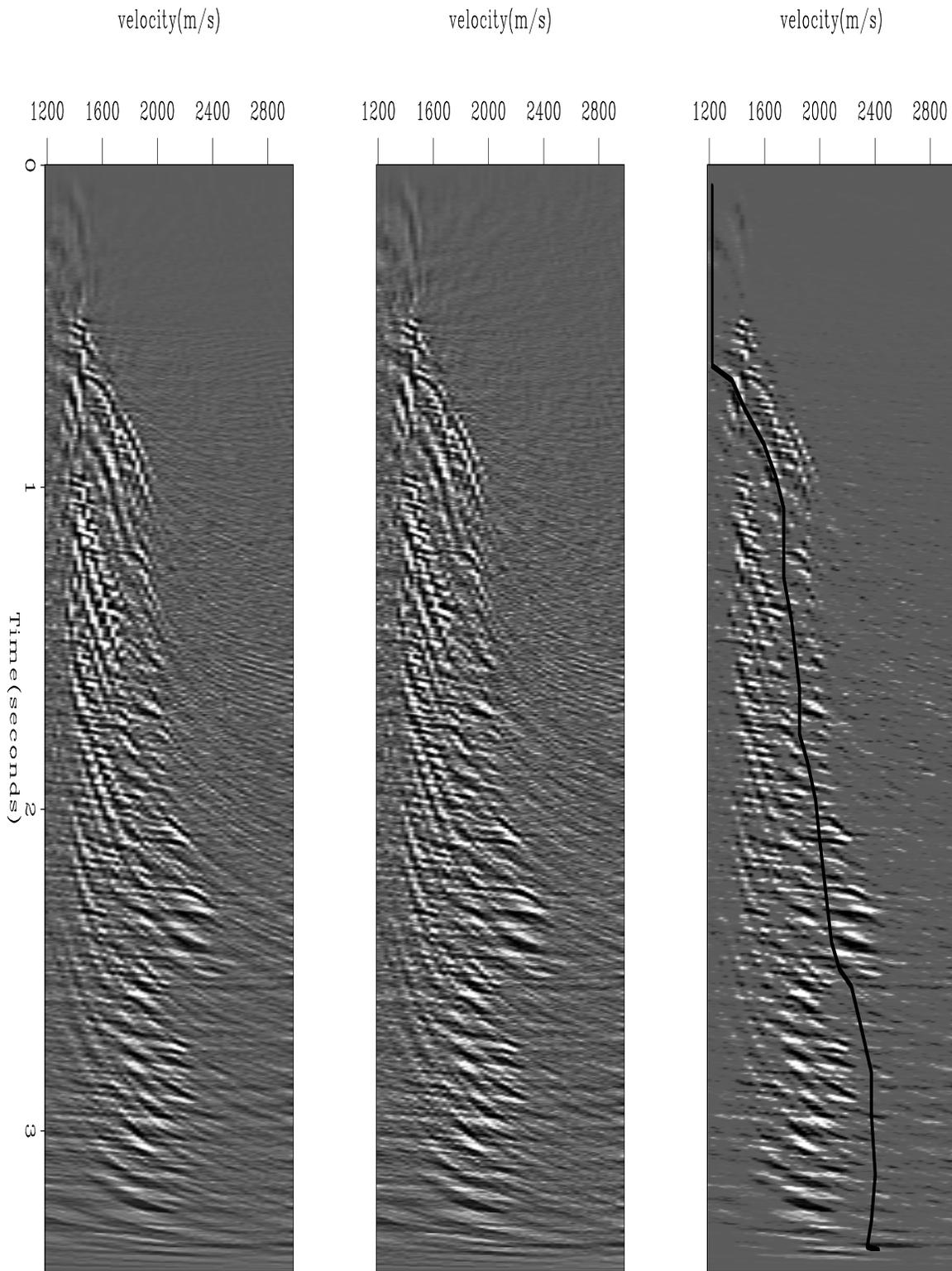


Figure 6: Left: l^2 model. Middle: l^1 model. Right: l^1 with l^1 regularization. The line shows the limit of the muting process that separates “guessed” multiples on the left from “guessed” primaries on the right for the spiky model. `antoine1-comp_scan` [CR]

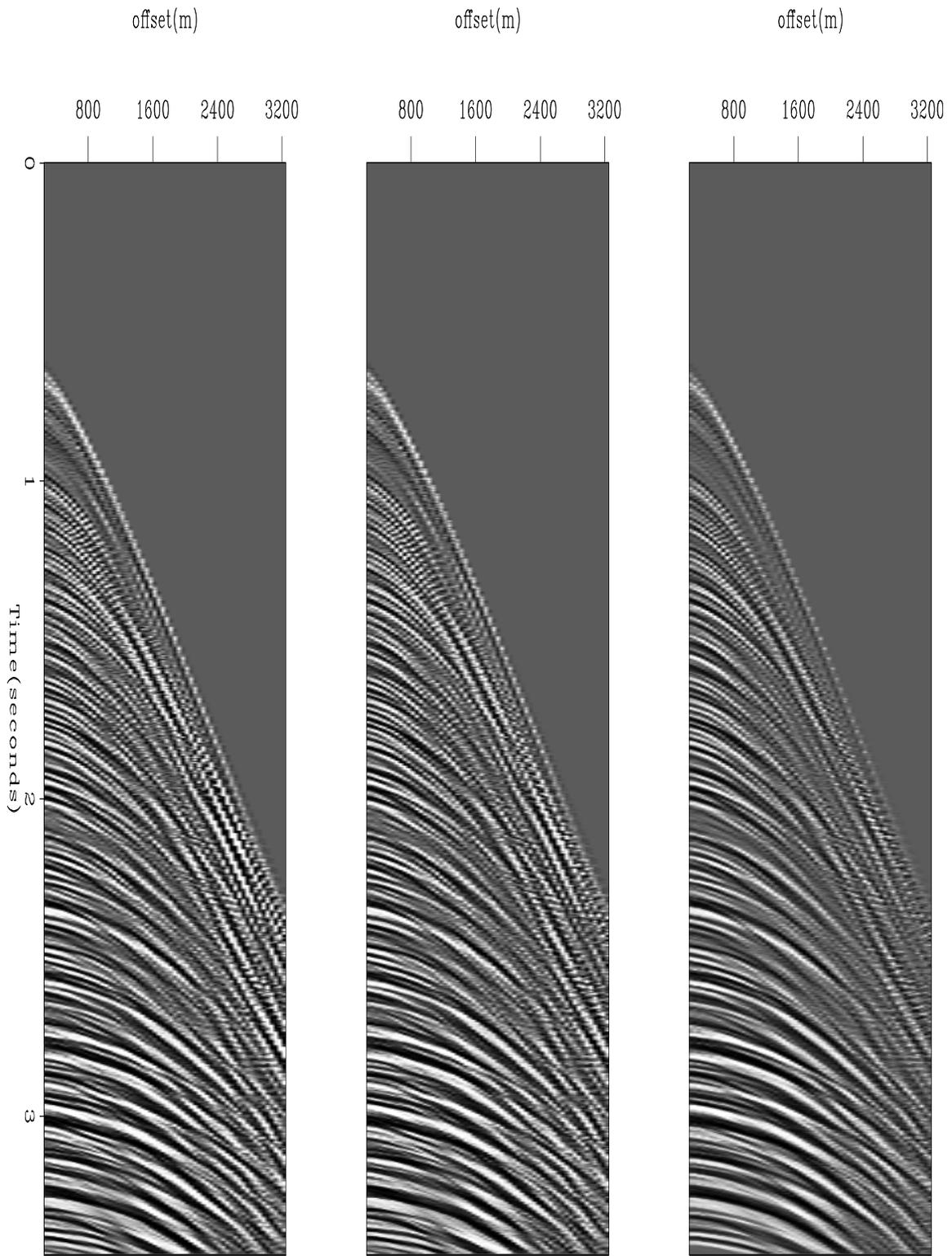


Figure 7: Predicted multiples. Left: l^2 multiples. Middle: l^1 multiples. Right: l^1 with l^1 regularization multiples. `antoine1-comp_mult` [CR]

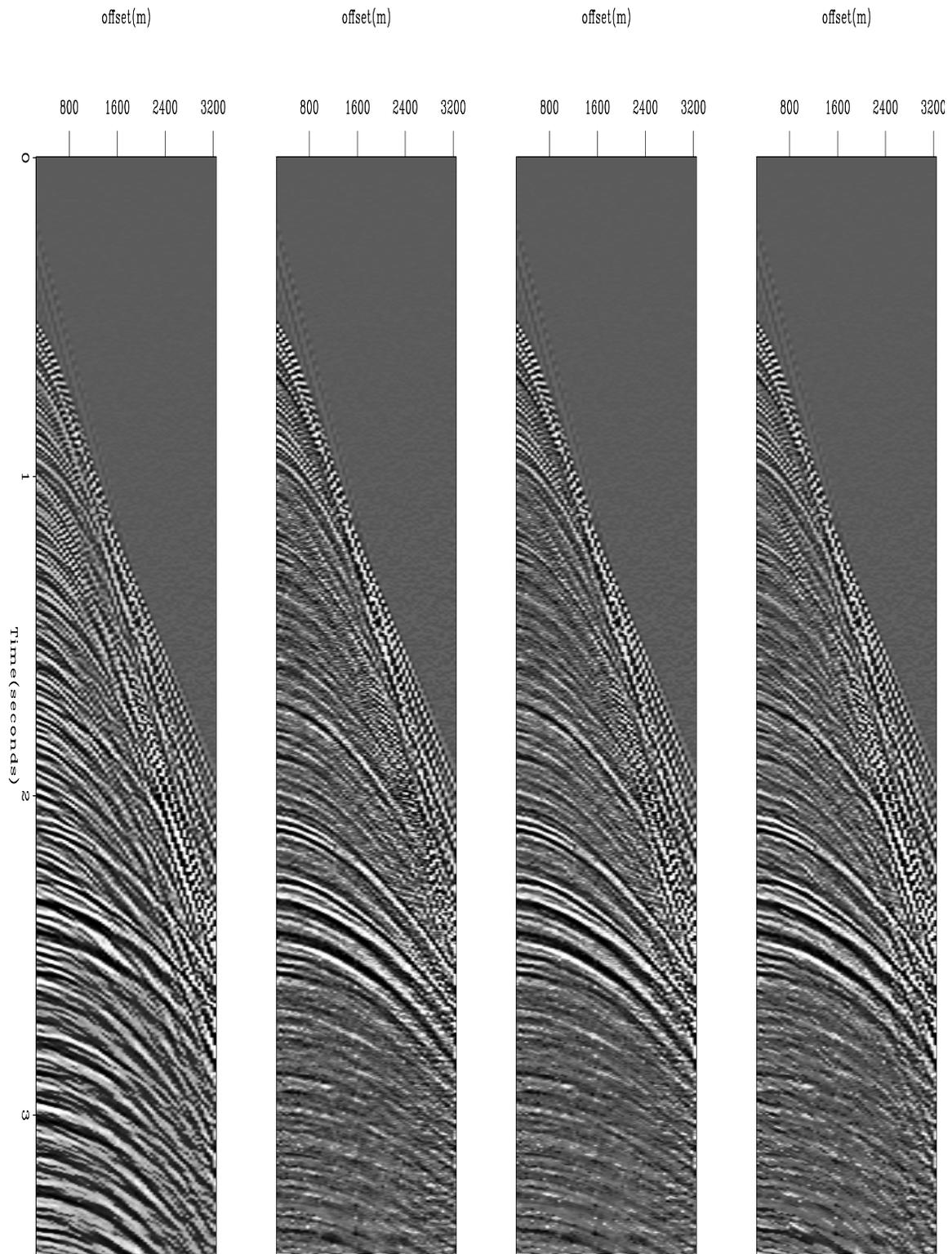


Figure 8: Gathers after multiple attenuation. Left: input data with multiples. Middle-left: l^2 multiple attenuation. Middle-right: l^1 multiple attenuation. Right: l^1 with l^1 regularization multiple attenuation. [antoine1-comp](#) [CR]

there are no noticeable differences between the three methods in the stacked sections. In addition, as is often the case, the stacking is such a powerful multiple-suppression method that it attenuates multiples sufficiently well: the simple stack of the input data with multiples looks fairly close to the stacked section without multiples (see Figure 9).

Discussion

The overall results may seem disappointing: the simple stacking of the input data give similar results to the stacking of multiple-free gathers using l^2 , l^1 and l^1 with l^1 regularization. Nonetheless, data with more complex multiples would certainly lead to different conclusions. In addition, these data demonstrate that when the gathers are not particularly noisy, which is the case here, the l^1 norm and the l^2 norm behave similarly. The l^1 norm with l^1 regularization produces expected sparse velocity panels with very bad convergence properties, meaning that we need to think about new strategies to improve it. Claerbout (2000, Personal communication) recently suggested that I minimize

$$f(\mathbf{m}) = \|\mathbf{R}(\mathbf{H}\mathbf{m} - \mathbf{d})\|$$

where \mathbf{R} is the Prediction Error Filter (PEF) of the residual. The PEF would help to obtain Independent Identically Distributed (IID) variables in the residual. Another idea is to minimize

$$f(\mathbf{m}) = |\mathbf{F} \cdot \mathbf{T}_{2-D}(\mathbf{H}\mathbf{m} - \mathbf{d})|_{\text{Huber}}$$

where $F \cdot T_{2-D}$ is the 2-D Fourier Transform. The idea behind this last equation is that the far offset data, which the l^1 norm with l^1 regularization does not fit very well, creates (almost) mono-frequency patterns in the residual that would map in a very localized area of the Fourier space with high amplitudes (two points for a perfectly mono-frequency event with one slope). The Huber norm with an appropriate threshold would treat these focused energies as outliers and get rid of them.

CONCLUSION

I have shown that the multiple reflections may be attenuated in the prestack domain using the Hyperbolic Radon Transform. Different inverse problems have been solved to obtain the velocity panels showing different properties: (1) the l^2 and l^1 norm produce comparable results in the velocity space and for the reconstructed data, (2) the l^1 norm with l^1 regularization shows spikier results in the velocity domain but converge much more slowly than the other methods, and (3) the stacked sections of the multiple-free data are very similar for the different inverse problems. The good quality of the data explains the small discrepancy between the l^2 and the l^1 norm; more noisy gathers would lead to different conclusions.

REFERENCES

Brown, M., Clapp, B., and Marfurt, K., 1999, Predictive signal/noise separation of ground-roll contaminated data: SEP-102, 111-128.

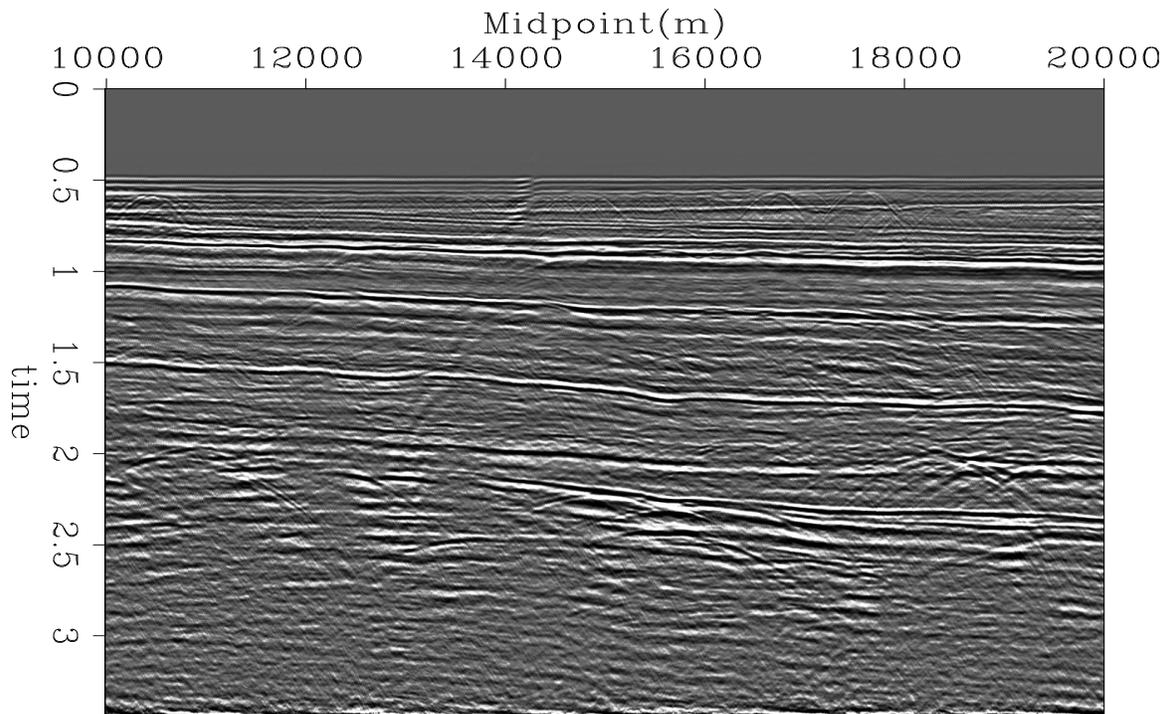


Figure 9: Stacked section of the input data with multiples. `antoine1-stack_IN` [CR]

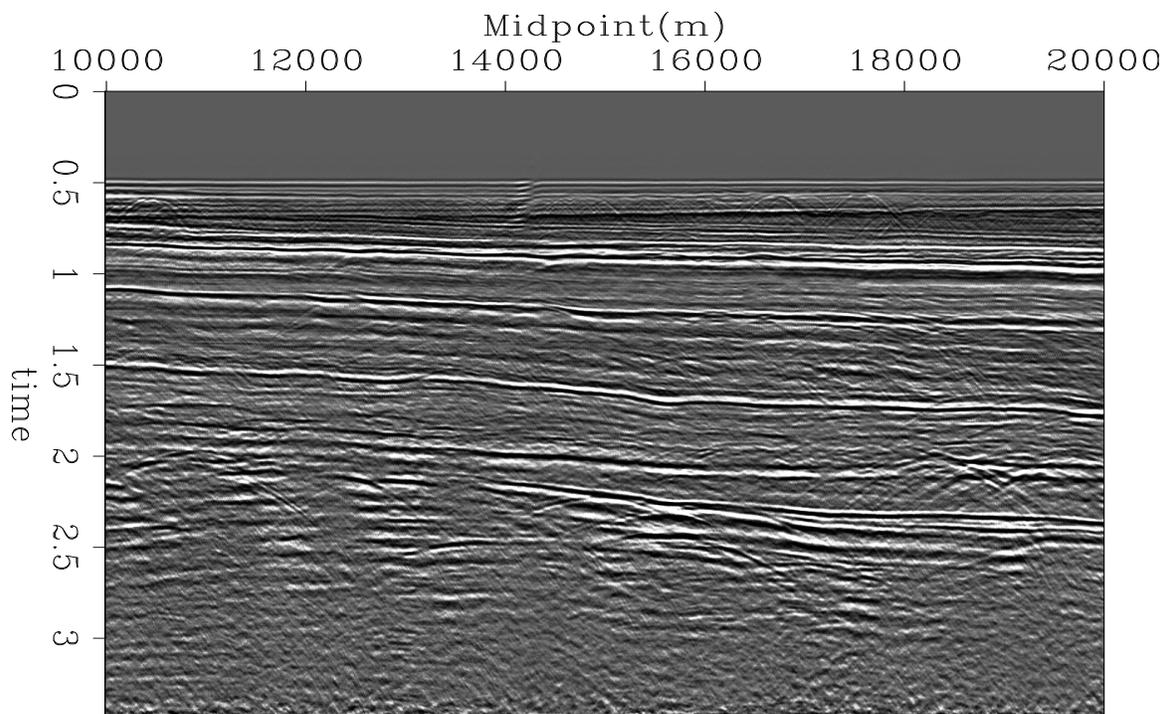


Figure 10: Stacked section after multiple suppression using the l^2 norm. `antoine1-stack_L2` [CR]

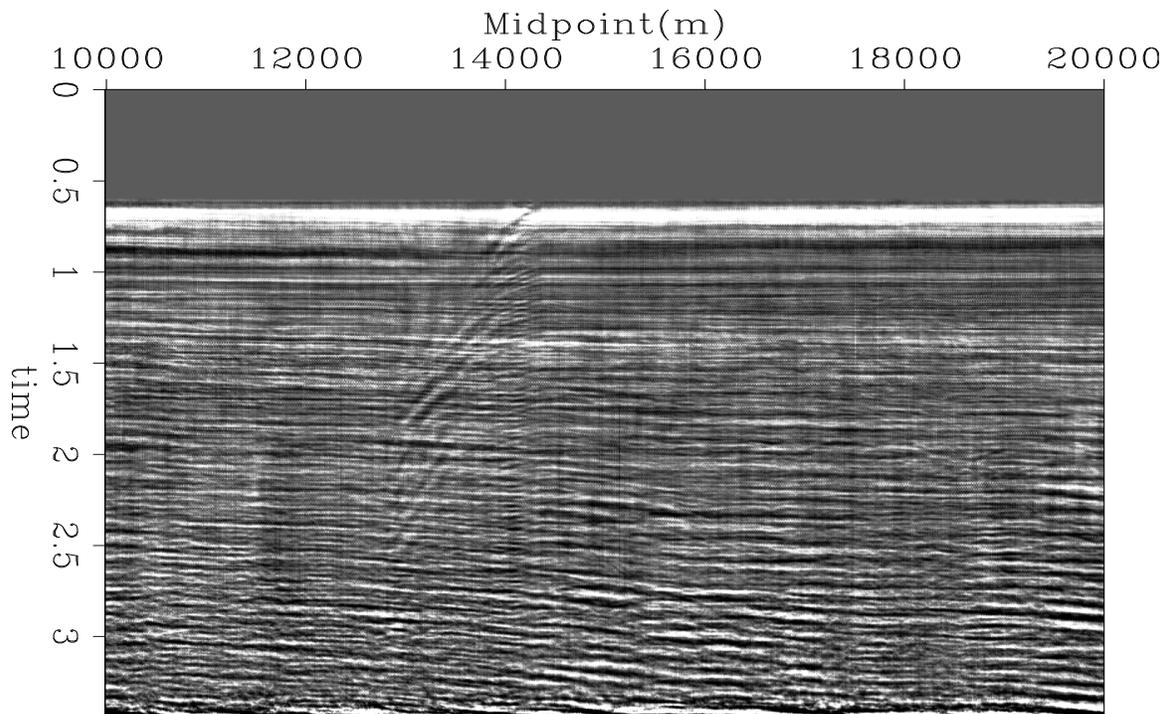


Figure 11: Difference between the stacked section with multiples and the stacked section without multiples using the l^2 norm. `antoine1-comp_stack_L2` [CR]

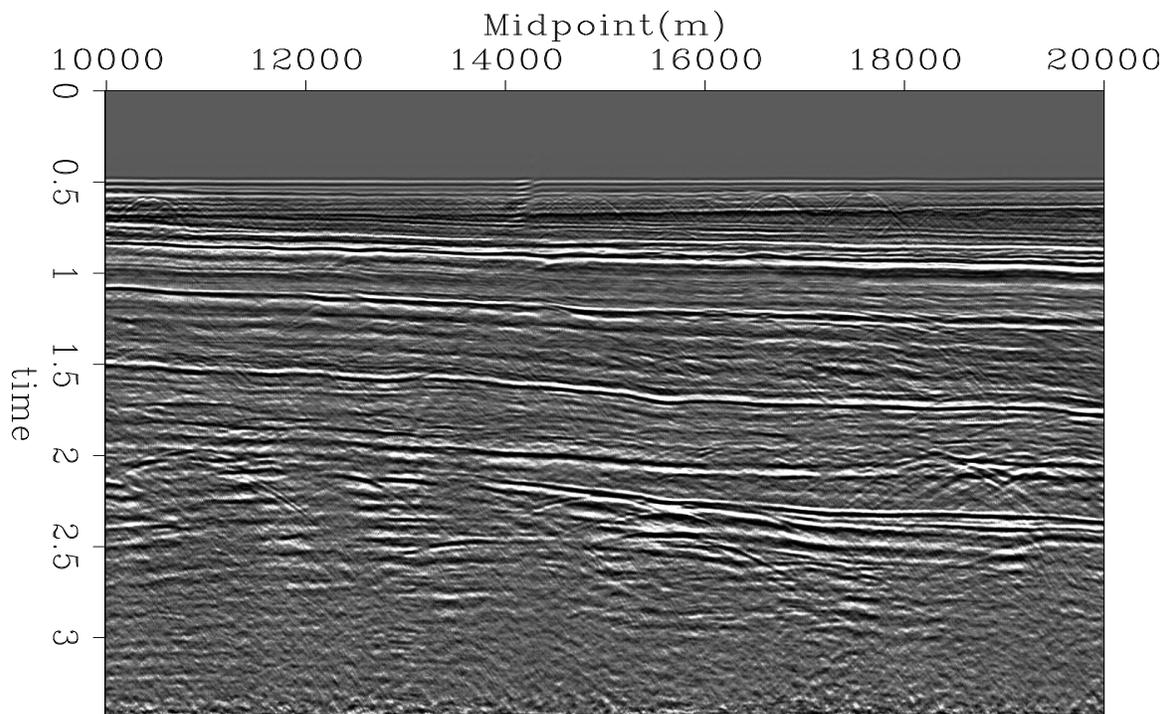


Figure 12: Stacked section after multiple suppression using the l^1 norm. `antoine1-stack_L1` [CR]

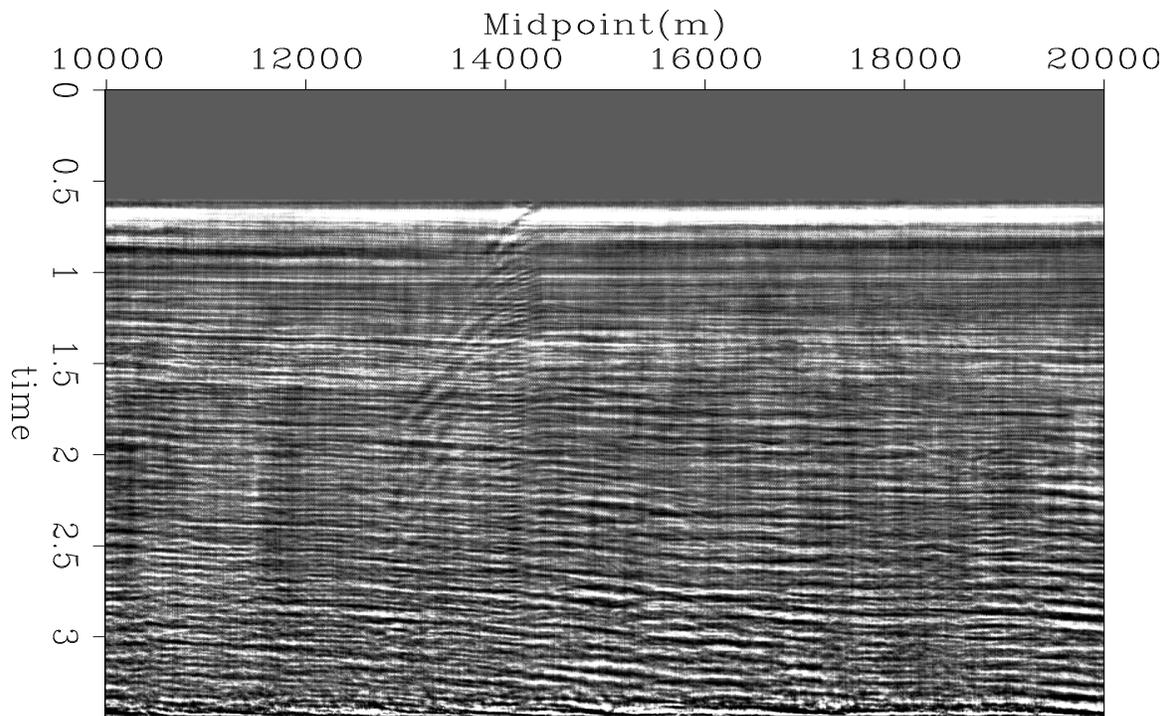


Figure 13: Difference between the stacked section with multiples and the stacked section without multiples using the l^1 norm. `antoine1-comp_stack_L1` [CR]

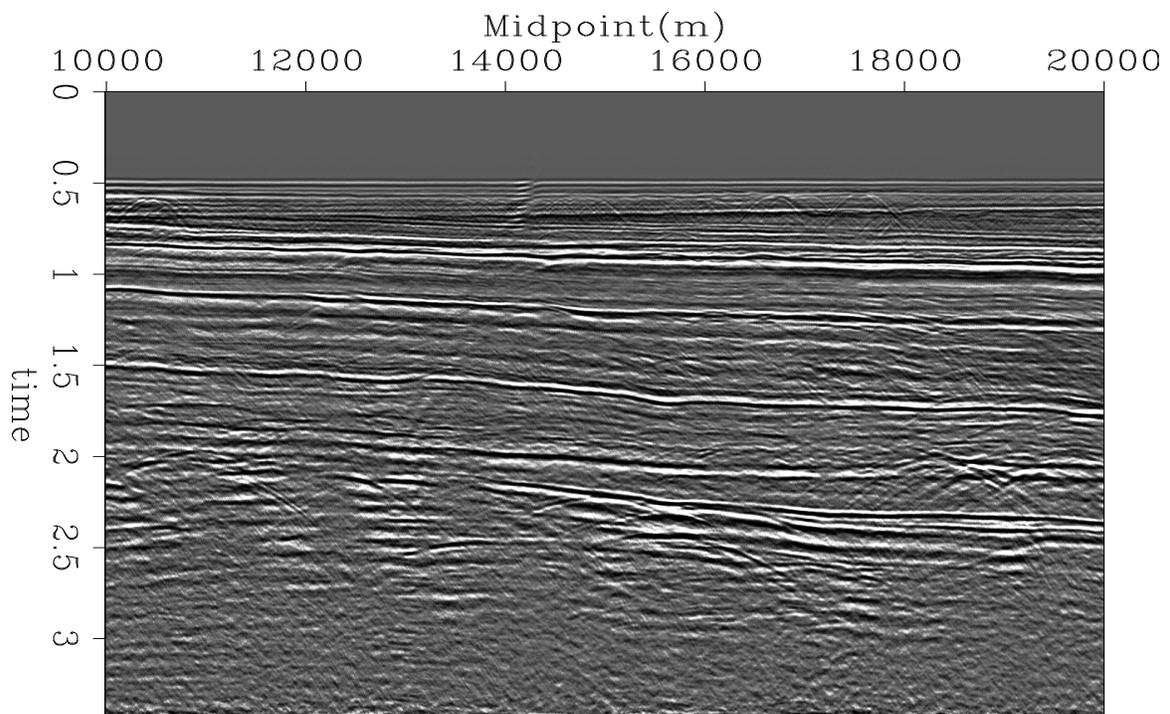


Figure 14: Stacked section after multiple suppression using the l^1 norm with l^1 regularization. `antoine1-stack_L1L1` [CR]

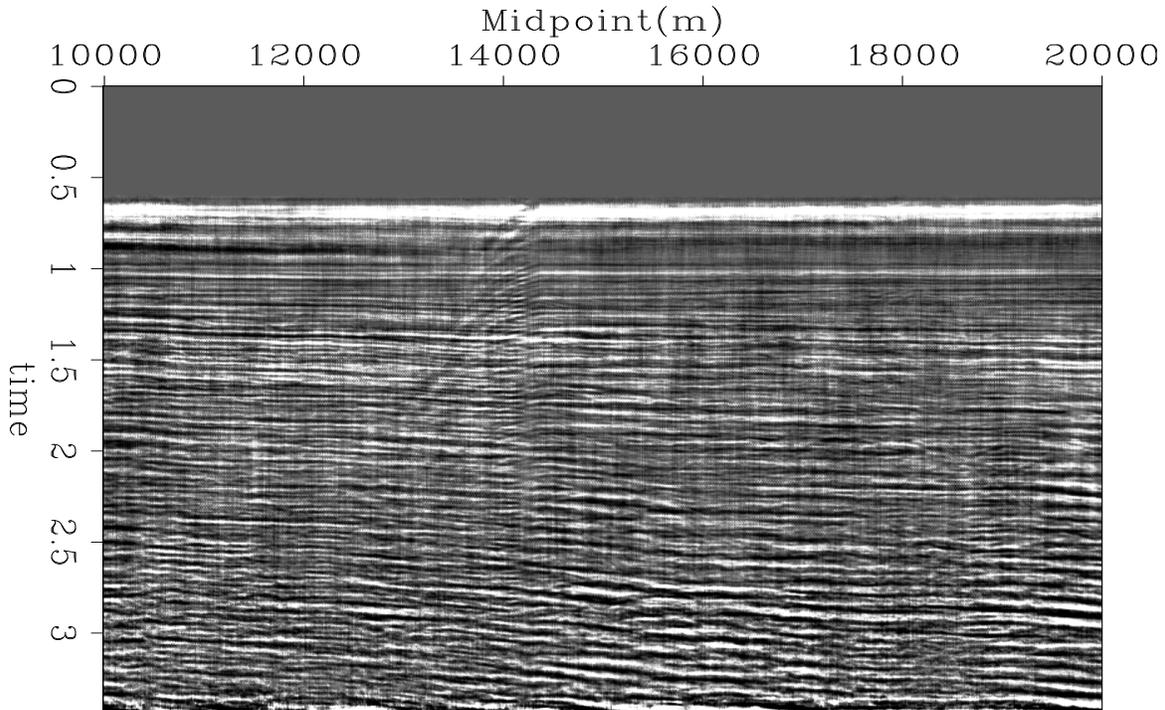


Figure 15: Difference between the stacked section with multiples and the stacked section without multiples using the l^1 norm with l^1 regularization. `antoine1-comp_stack_L1L1` [CR]

Bube, K. P., and Langan, R. T., 1997, Hybrid λ_1/λ_2 minimization with applications to tomography: *Geophysics*, **62**, no. 04, 1183–1195.

Claerbout, J. F., and Black, J. L., 1997, Basic Earth Imaging: Class notes, <http://sepwww.stanford.edu/sep/prof/index.html>.

Claerbout, J. F., and Muir, F., 1973, Robust modeling with erratic data: *Geophysics*, **38**, 820–844.

Claerbout, J., 1996, Conjugate-direction Huber regression: *SEP-92*, 229–235.

Clapp, B., and Brown, M., 1999, Applying SEP's latest tricks to the multiple suppression problem: *SEP-102*, 91–100.

Doicin, D., and Spitz, S., 1991, Multichannel extraction of water bottom peg legs pertaining to a high amplitude reflection: 61th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1439–1442.

Dragoset, B., and MacKay, S., 1993, Surface multiple attenuation and subsalt imaging: 63rd Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1099–1102.

Fomel, S., and Claerbout, J. F., 1995, Searching the sea of galilee: *SEP-94*, 259–270.

- Guitton, A., and Symes, W. W., 1999, Robust and stable velocity analysis using the Huber function: 69th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1166–1169.
- Guitton, A., 2000a, Huber solver versus IRLS algorithm for quasi L1 inversion: SEP-103, 255–271.
- Guitton, A., 2000b, Implementation of a nonlinear solver for minimizing the Huber norm: SEP-103, 281–289.
- Huber, P. J., 1973, Robust regression: Asymptotics, conjectures, and Monte Carlo: Ann. Statist., **1**, 799–821.
- Kabir, N. M. M., and Marfurt, K. J., 1999, Toward true amplitude multiple removal: The Leading Edge, **18**, 66–73.
- Kelley, C. T., 1999, Iterative methods for optimization: SIAM in applied mathematics.
- Liu, D. C., and Nocedal, J., 1989, On the limited memory BFGS method for large-scale optimization: Mathematical Programming, **45**, 503–528.
- Lumley, D., Nichols, D., Ecker, C., Rekdal, T., and Berlioux, A., 1995, Amplitude-preserved processing and analysis of the mobil avo data set: SEP-84, 125–152.
- More, J. J., and Thuente, J., 1994, Line search algorithms with guaranteed sufficient decrease: ACM Transactions on Mathematical Software, **20**, 286–307.
- Nichols, D., 1994, Velocity-stack inversion using Lp norms: SEP-82, 1–16.
- Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, **95**, 339–353.
- Sacchi, M. D., and Ulrych, T. J., 1995, High-resolution velocity gathers and offset space reconstruction: Geophysics, **60**, no. 04, 1169–1177.
- Spitz, S., 1999, Pattern recognition, spatial predictability, and subtraction of multiple events: The Leading Edge, **18**, 55–58.
- Taner, M. T., and Koehler, F., 1969, Velocity spectra-digital computer derivation and applications of velocity functions: Geophysics, **34**, no. 06, 859–881.
- Tarantola, A., 1987, Inverse Problem Theory: Elsevier Science Publisher.
- Thorson, J. R., and Claerbout, J. F., 1985, Velocity stack and slant stochastic inversion: Geophysics, **50**, no. 12, 2727–2741.
- Verschuur, D. J., Berkhout, A. J., and Wapenaar, C. P. A., 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, 1166–1177.

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

Weglein, A. B., 1999, Multiple attenuation: an overview of recent advances and the road ahead (1999): *The Leading Edge*, **18**, 40–44.

$(t - x)$ domain, pattern-based multiple separation

Robert G. Clapp and Morgan Brown¹

ABSTRACT

Pattern-based signal/noise separation is a common technique to suppress multiples. It can be formulated in the $t - x$ domain using non-stationary Prediction Error Filters (PEF). One can obtain a kinematically correct model of the multiples by downward continuation. The CMP gather and the corresponding multiple estimate are characterized by a space varying PEF. After applying a simple separation technique one can obtain CMP gathers where the multiple energy is significantly attenuated. The method is applied to synthetic and 2-D field CMP gathers.

INTRODUCTION

Multiple suppression is one of the largest problems facing the seismic industry. One common technique are the family of approaches generally referred to as ‘model based’ (Berryhill and Kim, 1986; Wiggins, 1988). These methods work by first getting an estimate of the models through downward continuation (Berryhill and Kim, 1986), computing the first term of the Neuman series (Ikelle et al., 1997), or some other method. Next, the primaries are estimated through some type of filtering operation using the estimated multiples. Recently, the problem has been formulated as a signal-noise separation problem in the frequency domain (Spitz, 1999; Bednar and Neale, 1999). These methods operate in the $f - x$ domain with the limiting assumption that the data are time-stationary.

Until recently the signal-noise method proposed by Spitz (1999) could not be formulated in the time domain because it involves dividing by a filter describing the multiple. Claerbout (1998) discovered that multi-dimensional PEFs can be mapped into 1-D, therefore making it possible to do inverse filtering in the time domain. The stationarity assumption inherent in PEF estimation can be overcome by estimating non-stationary filters (Crawley et al., 1998). As a result, Spitz’s (1999) method can be formulated to work with time domain PEFs (Clapp and Brown, 1999).

In this paper we show how the time domain formulation of Spitz’s approach can effectively attenuate multiples. We apply the method to a 2-D synthetic dataset and show that it is effective in both simple and complex areas. We then apply it on a 2-D real CMP gather. We show that our technique is successful in the attenuating most of the multiple energy with little loss of primary energy.

¹**email:** bob@sep.Stanford.EDU,morgan@sep.Stanford.EDU

METHODOLOGY

Signal to noise separation

Consider the recorded data \mathbf{d} to be the simple superposition of “signal” \mathbf{s} , i.e., reflection events, and “noise” \mathbf{n} , i.e., multiples: $\mathbf{d} = \mathbf{s} + \mathbf{n}$. For the special case of uncorrelated signal and noise, the so-called *Wiener estimator* is a filter, which when applied to the data, yields an optimal (least-squares sense) estimate of the embedded signal (Castleman, 1996). The frequency response \mathbf{H} of this filter is

$$\mathbf{H} = \frac{\mathbf{P}_s}{\mathbf{P}_n + \mathbf{P}_s}, \quad (1)$$

where \mathbf{P}_s and \mathbf{P}_n are the signal and noise power spectra, respectively. Abma (1995) and Claerbout (1999) solved a constrained least squares problem to separate signal from spatially uncorrelated noise:

$$\begin{aligned} \mathbf{Nn} &\approx 0 \\ \epsilon \mathbf{Ss} &\approx 0 \end{aligned} \quad (2)$$

subject to $\leftrightarrow \mathbf{d} = \mathbf{s} + \mathbf{n}$

where the operators \mathbf{N} and \mathbf{S} represent $t - x$ domain convolution with non-stationary PEF which whiten the unknown noise \mathbf{n} and signal \mathbf{s} , respectively, and ϵ is a Lagrange multiplier. Minimizing the quadratic objective function suggested by equation (2) with respect to \mathbf{s} leads to the following expression for the estimated signal:

$$\hat{\mathbf{s}} = (\mathbf{N}^T \mathbf{N} + \epsilon^2 \mathbf{S}^T \mathbf{S})^{-1} \mathbf{N}^T \mathbf{N} \mathbf{d} \quad (3)$$

By construction, the frequency response of a PEF approximates the inverse power spectrum of the data from which it was estimated. Thus, we see that the approach of equation (2) is similar to the Wiener reconstruction process. Spitz (1999) showed that for uncorrelated signal and noise, the signal can be expressed in terms of a PEF, \mathbf{D} , estimated from the data \mathbf{d} , and a PEF, \mathbf{N} , estimated from the noise model:

$$\mathbf{S} = \mathbf{DN}^{-1}. \quad (4)$$

Spitz' result applies to one-dimensional PEF's in the $f - x$ domain, but our use of the helix transform (Claerbout, 1998) permits stable inverse filtering with multidimensional $t - x$ domain filters. Substituting $\mathbf{S} = \mathbf{DN}^{-1}$ and applying the constraint $\mathbf{d} = \mathbf{s} + \mathbf{n}$ to equation (2) gives

$$\begin{aligned} \mathbf{Ns} &\approx \mathbf{Nd} \\ \epsilon \mathbf{DN}^{-1} \mathbf{s} &\approx \mathbf{0}. \end{aligned} \quad (5)$$

Iterative solutions to least-squares problems converge faster if the data and the model being estimated are both uncorrelated. To precondition this problem, we again appeal to the Helix

transform to make the change of variables $\mathbf{x} = \mathbf{S}\mathbf{s} = \mathbf{D}\mathbf{N}^{-1}\mathbf{s}$ or $\mathbf{s} = \mathbf{N}\mathbf{D}^{-1}\mathbf{x}$ and apply it to equation (5):

$$\begin{aligned} \mathbf{N}\mathbf{N}\mathbf{D}^{-1}\mathbf{x} &\approx \mathbf{N}\mathbf{d} \\ \epsilon\mathbf{x} &\approx \mathbf{0} \end{aligned} \tag{6}$$

After solving equation (6) for the preconditioned solution \mathbf{x} , we obtain the estimated signal by reversing the change of variables: $\hat{\mathbf{s}} = \mathbf{N}\mathbf{D}^{-1}\mathbf{x}$.

Filter estimation

A PEF (**a**) by definition is the filter that minimizes the energy when convolved with the data (**d**). To estimate a space-invariant filter, this amounts to applying the fitting goal,

$$\mathbf{0} \approx \mathbf{D}\mathbf{a}. \tag{7}$$

When estimating a space-varying PEF, the number of filter coefficients can quickly become more than the number of data points, creating an underdetermined problem. Crawley et al. (1999) proposed estimating space varying filter with radial patch. The fitting goals become

$$\begin{aligned} \mathbf{0} &\approx \mathbf{D}\mathbf{A}^{-1}\mathbf{p} \\ \mathbf{0} &\approx \epsilon\mathbf{A}^{-1}, \end{aligned} \tag{8}$$

where \mathbf{A}^{-1} is a preconditioning operator that smoothes in a radial direction (assuming that dips will be more consistent along radial lines, Figure 1.

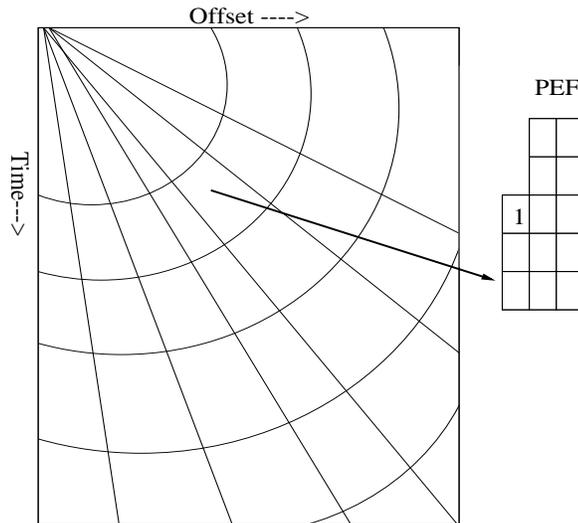


Figure 1: Space varying filter composition. A different filter is placed inside each micro-patch. The filter estimation problem is done globally, with the filter coefficient smoothed in a radial direction. `bob2-pef` [NR]

EXAMPLE

Synthetic example

To test the method, we chose an elastic synthetic data generated by BP. Figure 2 is the p-wave velocity model used to construct the data. Within the model, multiple behavior ranged from rather simple (away from the salt body) to much more complex (along the edge and under the salt). We then modeled the multiples by doing frequency Kirchoff upward continuation of both the sources and receivers to the sea-floor. For our first test we chose a CMP gather at 1 kft, away from the salt edge. The left panel of Figure 3 shows the original CMP gather. The center panel shows the upward continued gather at the same location. The right panel shows the predicted signal for the CMP gather. Note that we have done a good job eliminating the multiples with little loss of primary energy. Our second test was more of a challenge. We chose

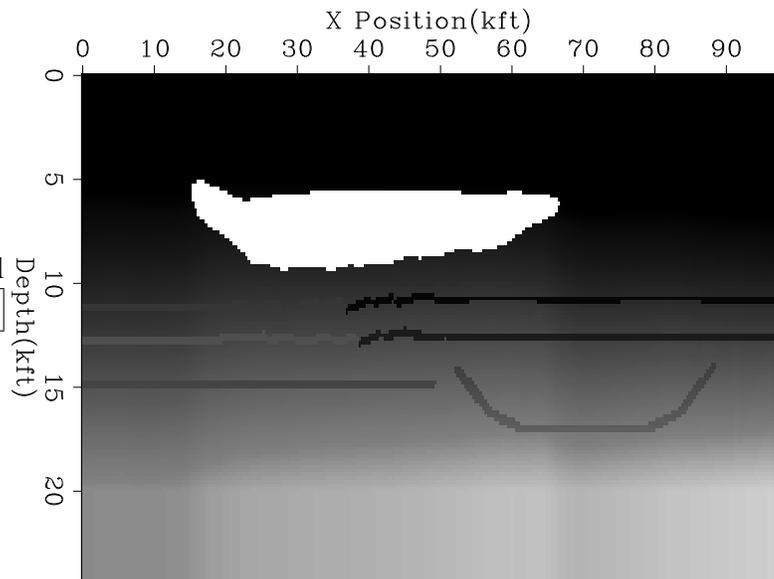


Figure 2: Synthetic velocity model for the BP dataset. [ER]

a CMP through the salt at 44 kft (Figure 4). Note that both the primary and multiple energy is significantly more complicated than in the previous example. The right panel of Figure 4 again shows our primary estimate. In this case our primary estimate isn't quite as good as Figure 3 but we still have done an acceptable job reducing the multiple energy.

Real data

For the real data example, we chose a dataset provided by Mobil. The data previously was by Lumley et al. (1994) using a hyperbolic radon technique and by Guitton (2000) with L1 hyperbolic multiple attenuation scheme. We found the water bottom and upward continued the data. The left panel of Figure 5 is a CMP gather from the data and the middle panel is our multiple estimate. If we look at our primary estimate (the right panel of Figure 5) we can see that we have removed a little primary energy and have little multiple energy in the gather, but overall, we have done an acceptable job attenuating the multiples.

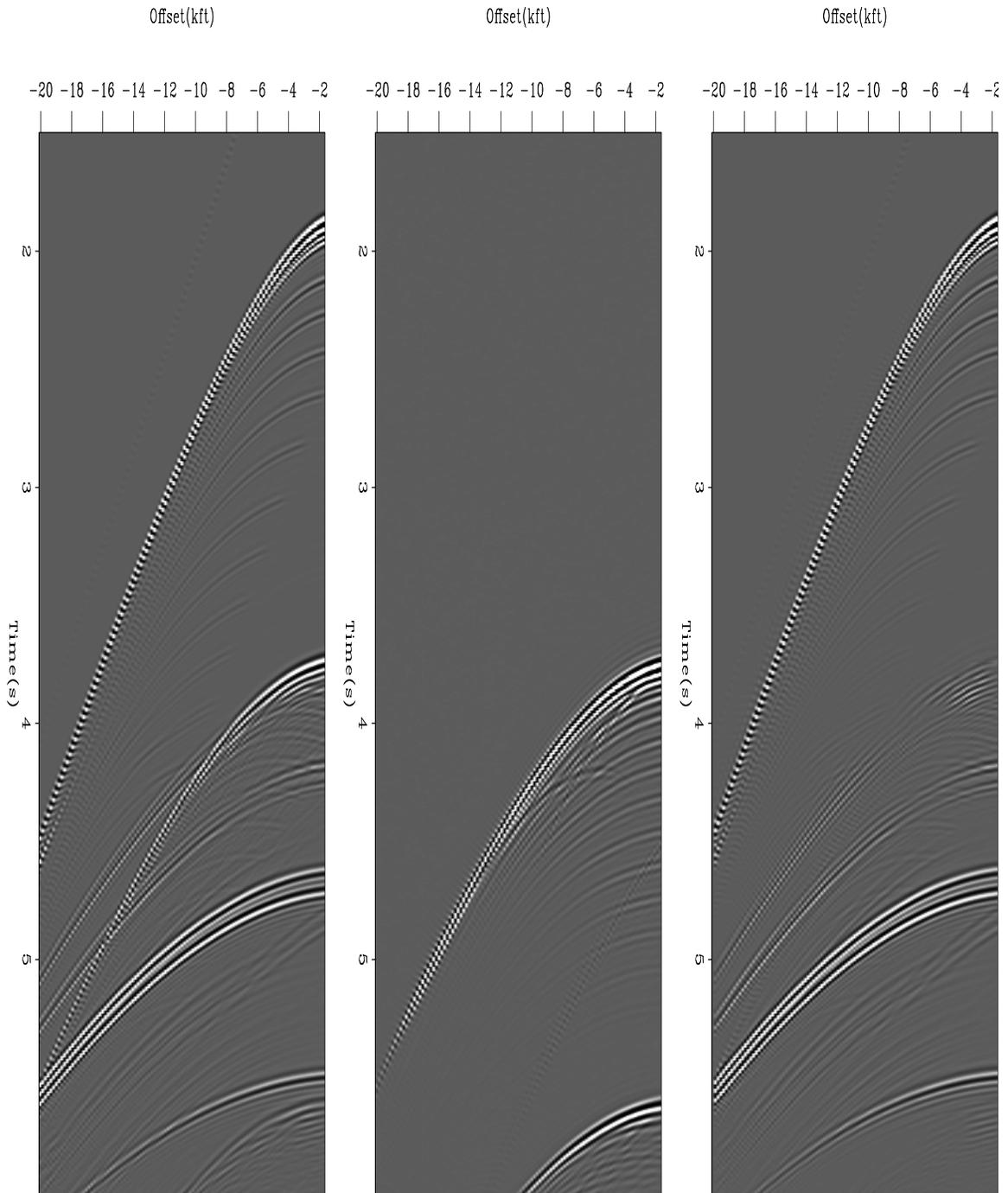


Figure 3: CMP gather at one kft. The left panel is the original CMP gather, the middle is the multiple estimate, and the right panel is our signal estimate. `bob2-bp-cmp1` [ER]

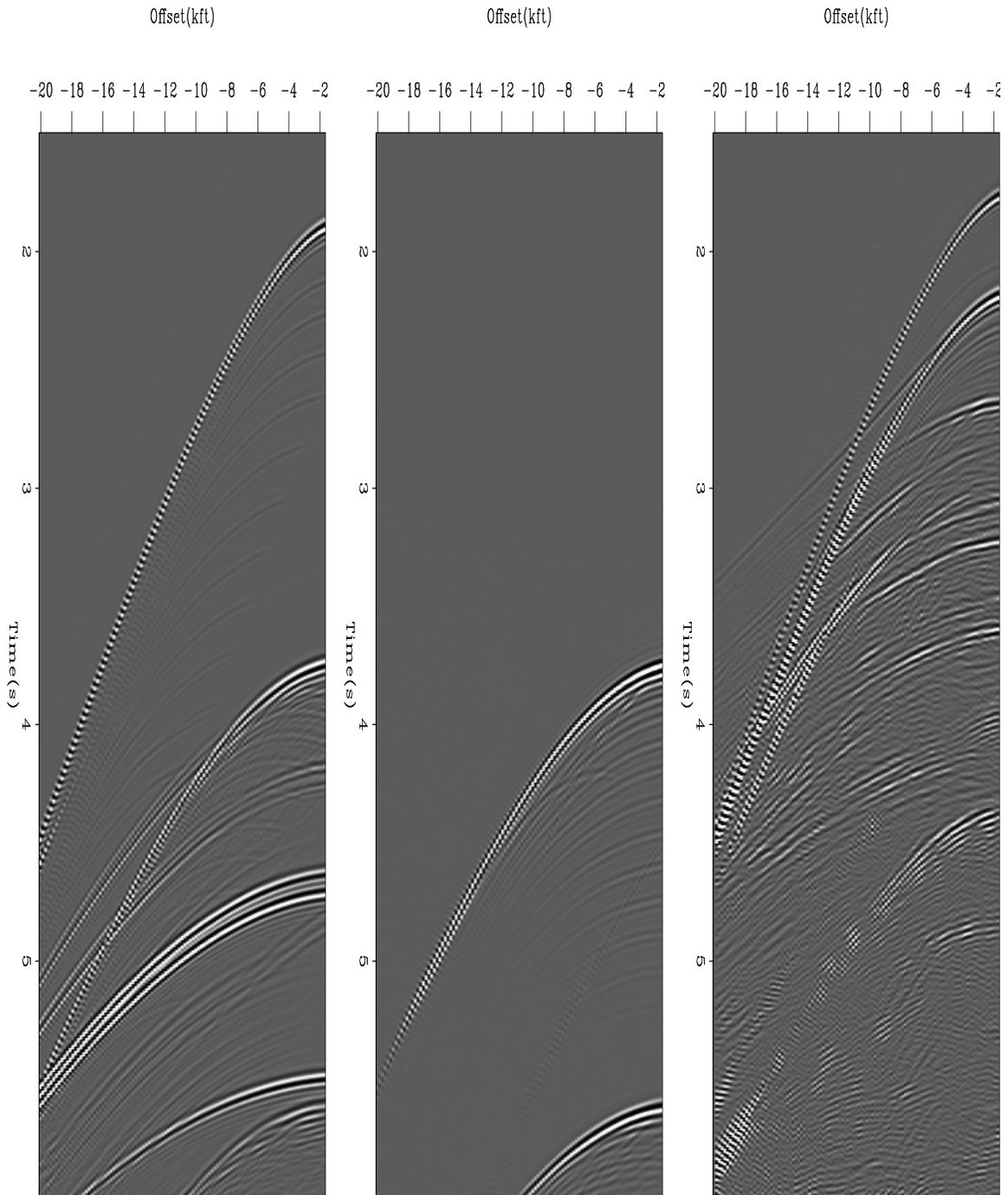


Figure 4: CMP gather at 44 kft. The left panel is the original CMP gather, the middle is the multiple estimate, and the right panel is our signal estimate. `bob2-bp-cmp2` [ER]

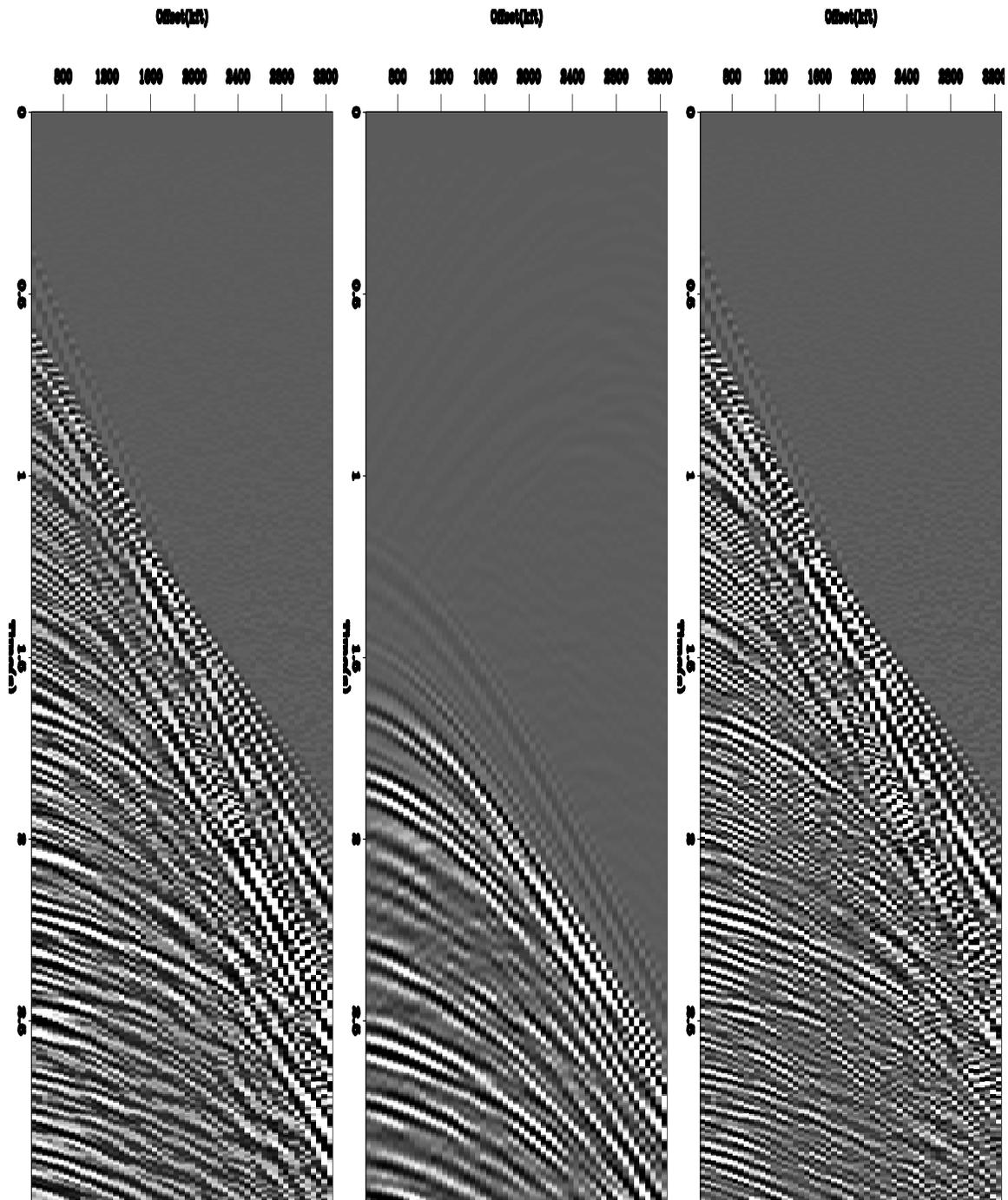


Figure 5: Multiple suppression result on real data. The left panel is the original CMP gather, the middle is the multiple estimate, and the right panel is our signal estimate. `bob2-mobil-cmp2` [ER]

PROBLEMS

Currently, the major weakness of this approach is its sensitivity to parameter choice. The separation fitting goals (6) apply the inverse of a non-stationary PEF. If that PEF isn't stable, the separation of the multiples and primaries is not possible. To get a stable filter we can increase ϵ in our filter estimation (9). Unfortunately, increasing ϵ decreases the quality of our prediction. By changing the size of our micro-patches, we can usually get a stable filter while obtaining a good prediction. At this stage we haven't figured an algorithm that can automatically change micro-patch size to obtain the desired combination, a stable non-stationary PEF that can satisfactorily predict the data.

CONCLUSIONS

On early tests, the separation technique was successful in suppressing multiples. Until the method can be made more stable, it can not be used in a production environment.

ACKNOWLEDGMENTS

We would like to thank BP for providing the synthetic dataset and Mobil for providing the field data to SEP.

REFERENCES

- Abma, R., 1995, Least-squares separation of signal and noise with multidimensional filters: Ph.D. thesis, Stanford University.
- Bednar, J., and Neale, G., 1999, A comparison of pattern and series-based multiple suppression: 69th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1056–1059.
- Berryhill, J. R., and Kim, Y. C., 1986, Deep-water peglegs and multiples - emulation and suppression: *Geophysics*, **51**, no. 12, 2177–2184.
- Castleman, K. R., 1996, *Digital image processing*: Prentice-Hall.
- Claerbout, J. F., 1998, Multi-dimensional recursive filtering via the helix: *Geophysics*, **63**, no. 5, 1532–1541.
- Claerbout, J., 1999, *Geophysical estimation by example: Environmental soundings image enhancement*: Stanford Exploration Project, <http://sepwww.stanford.edu/sep/prof/>.
- Clapp, R. G., and Brown, M., 1999, Applying sep's latest tricks to the multiple suppression problem: *SEP-102*, 91–100.

- Crawley, S., Clapp, R., and Claerbout, J., 1998, Deconvolution and interpolation with nonstationary filters: *SEP-97*, 183–192.
- Crawley, S., Claerbout, J., and Clapp, R., 1999, Interpolation with smoothly nonstationary prediction-error filters: 69th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1154–1157.
- Guitton, A., 2000, Prestack multiple attenuation using the hyperbolic Radon transform: *SEP-103*, 181–201.
- Ikelle, L. T., Roberts, G., and Weglein, A. B., 1997, Source signature estimation based on the removal of first-order multiple: *Geophysics*, , no. 6, 1904–1920.
- Lumley, D., Nichols, D., and Rekdal, T., 1994, Amplitude-preserved multiple suppression: *SEP-82*, 25–45.
- Spitz, S., 1999, Pattern recognition, spatial predictability, and subtraction of multiple events: *The Leading Edge*, **18**, no. 1, 55–58.
- Wiggins, J. W., 1988, Attenuation of complex water-bottom multiples by wave equation-based prediction and subtraction: *Geophysics*, **53**, no. 12, 1527–1539.

Spitz makes a better assumption for the signal PEF

Jon Claerbout and Sergey Fomel¹

ABSTRACT

In real-world extraction of signal from data we are not given the needed signal prediction-error filter (PEF). Claerbout has taken S , the PEF of the signal, to be that of the data, $S \approx D$. Spitz takes it to be $S \approx D/N$. Where noises are highly predictable in time or space, Spitz gets significantly better results. Theoretically, a reason is that the essential character of a PEF is contained *where it is small*.

INTRODUCTION

Knowledge of signal spectrum and noise spectrum allows us to find filters for optimally separating data \mathbf{d} into two components, signal \mathbf{s} and noise \mathbf{n} (Claerbout, 1999). Actually, it is the inverses of these spectra which are required. In Claerbout's textbook example (1999) he estimates these inverse spectra by estimating prediction-error filters (PEFs) from the data. He estimates both a signal PEF and a noise PEF from the same data \mathbf{d} . A PEF based on data \mathbf{d} might be expected to be named the data PEF D , but Claerbout estimates two different PEFs from \mathbf{d} and calls them the signal PEF S and the noise PEF N . They differ by being estimated with different number of adjustable coefficients, one matching a signal model (two plane waves) having three positions on the space axis, the other matching a noise model having one position on the space axis.

Meanwhile, using a different approach, Spitz (1999) concludes that the signal, noise, and data inverse spectra should be related by $D = SN$. The conclusion we reach in this paper is that Claerbout's estimate of S is more appropriately an estimate of the data PEF D . To find the most appropriate S and N we should use both the "variable templates" idea of Claerbout and the $D \approx SN$ idea of Spitz. Here we first provide a straightforward derivation of the Spitz insight and then we show some experimental results.

¹**email:** claerbout@stanford.edu, sergey@sep.stanford.edu
This paper should have been in SEP-102 but tragedy intervened.

BASIC THEORY

Signal spectrum plus the noise spectrum gives the data spectrum. Since a prediction-error filter tends to the inverse of a spectrum we have

$$\frac{1}{\overline{DD}} = \frac{1}{\overline{SS}} + \frac{1}{\overline{NN}} \quad (1)$$

$$\frac{1}{\overline{DD}} = \frac{\overline{SS} + \overline{NN}}{\overline{SN} \overline{SN}} \quad (2)$$

or

$$\overline{DD} = \frac{\overline{SN} \overline{SN}}{\overline{SS} + \overline{NN}} \quad (3)$$

Now we are ready for the Spitz approximation. Spitz builds his applications upon the assumption that we can estimate D and N from suitable chunks of raw data. His result may be obtained from (3) by ignoring its denominator getting $D \approx SN$ or

$$S \approx D/N \quad (4)$$

Ignoring the denominator in equation (3), is not so terrible an approximation as it might seem. Remember that PEFs are important *where they are small* because they are used as weighting functions. Where weighting functions are small, solutions are expected to be large. Although Claerbout's assumption $S \approx D$ might be somewhat valid for signal and data *spectra*, it is much less valid for their *PEFs*. In practice, signal unpolluted with noise is usually not available. Even a very good chunk of data tends to yield a poor estimate of the signal PEF S because the holes in the signal spectrum are easily intruded with noise.

Obviously the major difference between $S \approx D$ and $S \approx D/N$ is where the noise is large. Thus it is for "organized and predictable" noises (small N) where we expect to see the main difference.

Theoretically, we need not make the Spitz approximation. We could solve (1) for S by spectral factorization. Although the S obtained would be more theoretically satisfying, there would be some practical disadvantages. Getting the signal spectrum by subtracting that of the noise from that of the data leaves the danger of a negative result (which explodes the factorization). Thus, maintaining spectral positivity would require extra care. All these extra burdens are avoided by making the Spitz approximation. All the more so in applications with continuously varying estimates.

SIGNAL AND NOISE SEPARATION

We assume that the data vector \mathbf{d} is composed of the signal and noise components \mathbf{s} and \mathbf{n} :

$$\mathbf{d} = \mathbf{s} + \mathbf{n} . \quad (5)$$

If both the signal and noise prediction-error filters S and N are known, then the signal can be extracted from the data by solving the following system by the least squares method:

$$\mathbf{0} \approx \mathbf{N}\mathbf{n} = \mathbf{N}(\mathbf{d} - \mathbf{s}); \quad (6)$$

$$\mathbf{0} \approx \epsilon \mathbf{S}\mathbf{s}, \quad (7)$$

where ϵ is a scalar scaling coefficient, reflecting the presumed signal-to-noise ration (Claerbout, 1999).

The formal solution of system (6-7) has the form of a *projection filter*:

$$\mathbf{s} = \left(\frac{\mathbf{N}'\mathbf{N}}{\mathbf{N}'\mathbf{N} + \epsilon^2 \mathbf{S}'\mathbf{S}} \right) \mathbf{d}. \quad (8)$$

Analogously, the signal vector is expressed as

$$\mathbf{n} = \mathbf{d} - \mathbf{s} = \left(\frac{\epsilon^2 \mathbf{S}'\mathbf{S}}{\mathbf{N}'\mathbf{N} + \epsilon^2 \mathbf{S}'\mathbf{S}} \right) \mathbf{d}. \quad (9)$$

In 1-D or $F-X$ setting, one can accomplish the division in formulas (8) and (9) directly by spectral factorization and inverse recursive filtering (Soubaras, 1995, 1994). A similar approach can be applied in the case of $T-X$ or $F-XY$ filtering with the help of the helix transform (Claerbout, 1998; Ozdemir et al., 1999) or by solving system (6-7) directly with an iterative method (Abma, 1995).

Claerbout's approach, implemented in the examples of *GEE* (Claerbout, 1999), is to estimate the signal and noise PEFs S and N from the data \mathbf{d} by specifying different shape templates for these two filters. The filter estimates can be iteratively refined after the initial signal and noise separation. In some examples, such as those shown in this paper, the signal and noise templates are not easily separated. When the signal template behaves as an extension of the noise template so that the shape of S completely embeds the shape of N , our estimate of S serves as a predictor of both signal and noise. We might as well consider it as D , the prediction-error filter for the data.

Spitz (1999) argues that the data PEF D can be regarded as the convolution of the signal and noise PEFs S and N . This assertion suggests the following algorithm:

1. Estimate D and N .
2. Estimate S by deconvolving (polynomial division) D by N .
3. Solve the least-square system (6-7).

To avoid the division step, we suggest a simple modification of Spitz's algorithm, which results from multiplying both equations in system (6-7) by the noise filtering operator \mathbf{N} . The resulting system has the form

$$\mathbf{0} \approx \mathbf{N}^2 \mathbf{n} = \mathbf{N}^2 (\mathbf{d} - \mathbf{s}); \quad (10)$$

$$\mathbf{0} \approx \epsilon \mathbf{N}\mathbf{S}\mathbf{s} = \epsilon \mathbf{D}\mathbf{s}. \quad (11)$$

The modified algorithm is

1. Estimate D and N .
2. Convolve N with itself.
3. Solve the least-square system (10-11).

The formal least-squares solution of system (10-11) is

$$\mathbf{s} = \left(\frac{\mathbf{N}'\mathbf{N}'\mathbf{N}\mathbf{N}}{\mathbf{N}'\mathbf{N}'\mathbf{N}\mathbf{N} + \epsilon^2\mathbf{D}'\mathbf{D}} \right) \mathbf{d} = \left(\frac{\mathbf{N}'\mathbf{N}'\mathbf{N}\mathbf{N}}{\mathbf{N}'\mathbf{N}'\mathbf{N}\mathbf{N} + \epsilon^2\mathbf{N}'\mathbf{S}'\mathbf{S}\mathbf{N}} \right) \mathbf{d}. \quad (12)$$

Comparing (12) with (8), we can see that both the numerator and the denominator in the two expressions differ by the same multiplier $\mathbf{N}'\mathbf{N}$. This multiplication should not effect the result of projection filtering.

Figure 1 shows a simple example of signal and noise separation taken from *GEE* (Claerbout, 1999). The signal consists of two crossing plane waves with random amplitudes, and the noise is spatially random. The data and noise T - X prediction-error filters were estimated from the same data by applying different filter templates. The template for D is

```

a a
a a
a a
1 a a
a a a
a a a
a a a

```

where the a symbol represents adjustable coefficients. The data filter shape has three columns, which allows it to predict two plane waves with different slopes. The noise filter N has only one column. Its template is

```

1
a
a
a

```

The noise PEF can estimate the temporal spectrum but would fail to capture the signal predictability in the space direction. Figure 2 shows the result of applying the modified Spitz method according to equations (10-11). Comparing figures 1 and 2, we can see that using a modified system of equations brings a slightly modified result with more noise in the signal but more signal in the noise. It is as if ϵ has changed, and indeed this could be the principal effect of neglecting the denominator in equation (3).

To illustrate a significantly different result using the Spitz insight we examine the new situation shown in Figures 3 and 4. The wave with the positive slope is considered to be

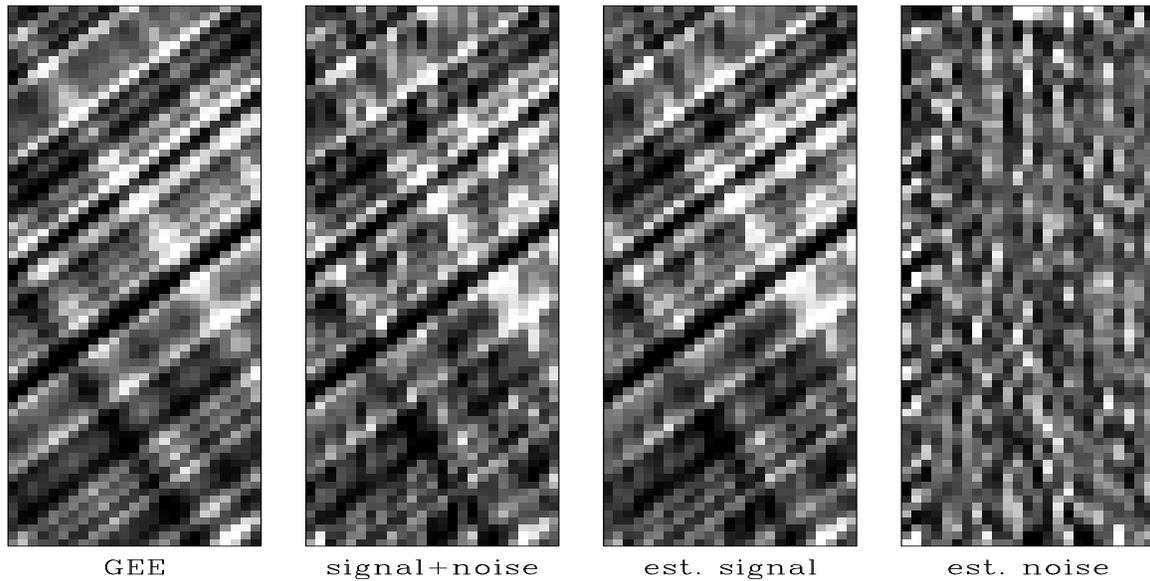


Figure 1: Signal and noise separation with the original GEE method. The input signal is on the left. Next is that signal with random noise added. Next are the estimated signal and the estimated noise. [jon1-signoi90](#) [ER]

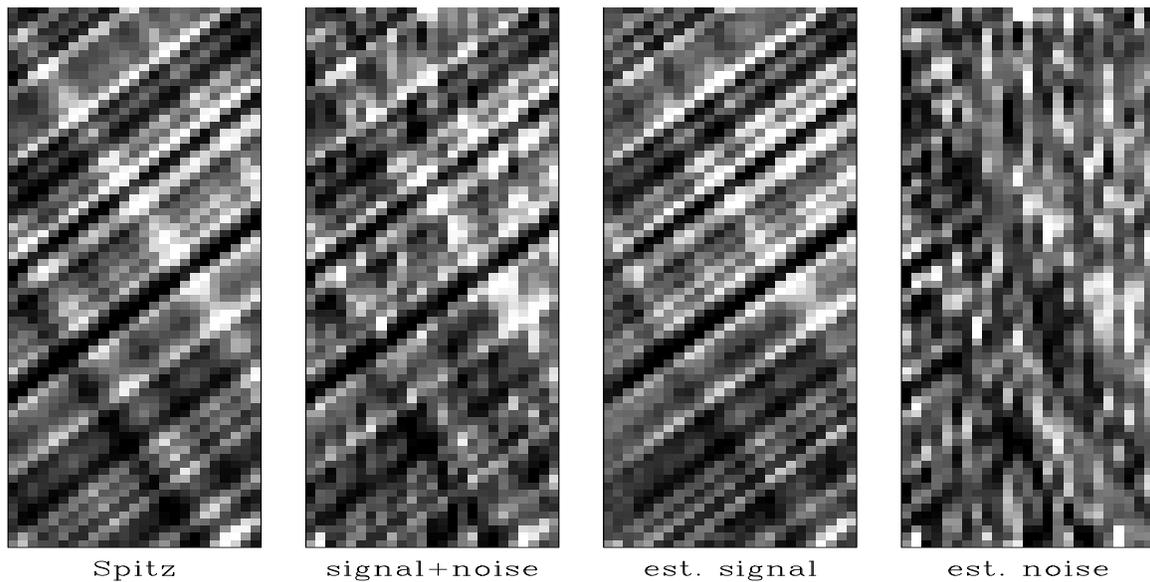


Figure 2: Signal and noise separation with the modified Spitz method. The input signal is on the left. Next is that signal with random noise added. Next are the estimated signal and the estimated noise. [jon1-signoi](#) [ER]

regular noise; the other wave is signal. The noise PEF N was estimated from the data by restricting the filter shape so that it could predict only positive slopes. The corresponding template is

```

a
1 a

```

The data PEF template is

```

a a
a a
1 a a
a a a
a a a

```

Using the data PEF as a substitute for the signal PEF produces a poor result, shown in Figure 3. We see a part of the signal sneaking into the noise estimate. Using the modified Spitz method, we obtain a clean separation of the plane waves (Figure 4).

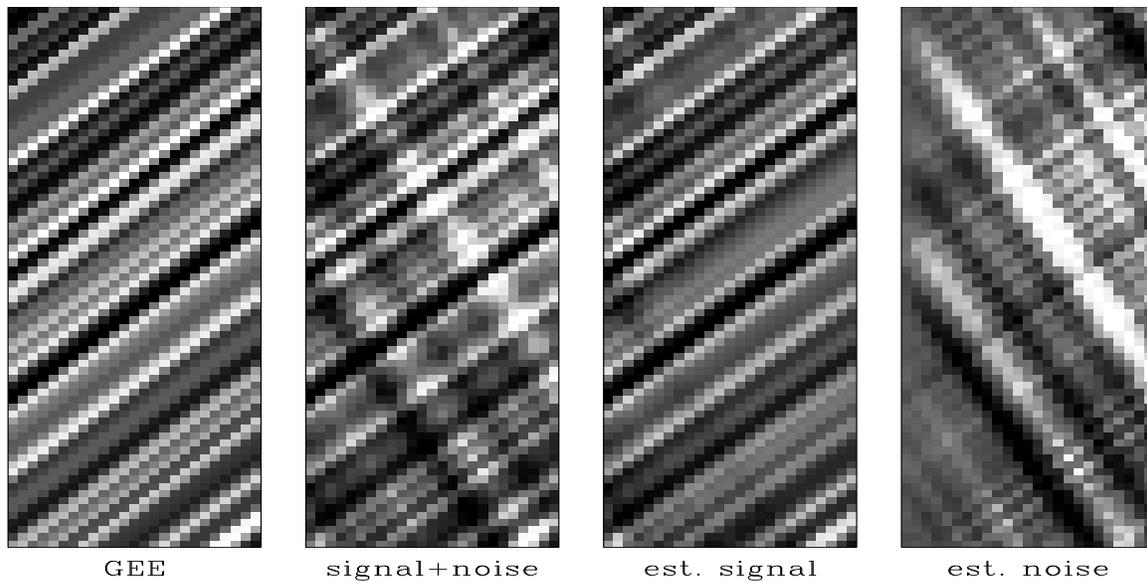


Figure 3: Plane wave separation with the GEE method. The input signal is on the left. Next is that signal with noise added. Next are the estimated signal and the estimated noise. [jon1-planes90](#) [ER]

Clapp and Brown (1999, 2000) and Brown et al. (1999) show applications of the least-squares signal-noise separation to multiple and ground-roll elimination.

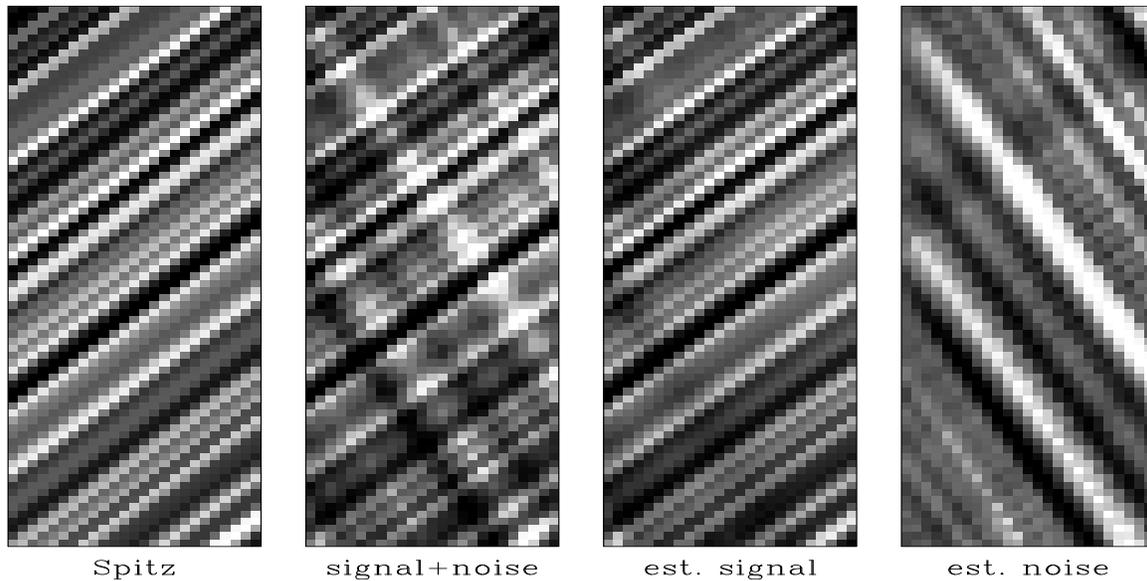


Figure 4: Plane wave separation with the modified Spitz method. The input signal is on the left. Next is that signal with noise added. Next are the estimated signal and the estimated noise. [jon1-planes](#) [ER]

ACKNOWLEDGMENTS

Conversations with our colleagues Bob Clapp and Morgan Brown led us to a better understanding of the Spitz approach.

REFERENCES

- Abma, R., 1995, Least-squares separation of signal and noise with multidimensional filters: Ph.D. thesis, Stanford University.
- Brown, M., Clapp, R. G., and Marfurt, K., 1999, Predictive signal/noise separation of groundroll-contaminated data: SEP-102, 111-128.
- Claerbout, J., 1998, Multidimensional recursive filters via a helix: Geophysics, **63**, 1532-1541.
- Claerbout, J., 1999, Geophysical estimation by example: Environmental soundings image enhancement: Stanford Exploration Project, <http://sepwww.stanford.edu/sep/prof/>.
- Clapp, R. G., and Brown, M., 1999, Applying sep's latest tricks to the multiple suppression problem: SEP-102, 91-100.
- Clapp, R. G., and Brown, M., 2000, ($t - x$) domain, pattern-based multiple separation: SEP-103, 201-210.

- Ozdemir, A., Ozbek, A., Ferber, R., and Zerouk, K., 1999, *F-XY* projection filtering using helical transformation: 69th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1231–1234.
- Soubaras, R., 1994, Signal-preserving random noise attenuation by the *F-X* projection: 64th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1576–1579.
- Soubaras, R., 1995, Prestack random and impulsive noise attenuation by *F-X* projection filtering: 65th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 711–714.
- Spitz, S., 1999, Pattern recognition, spatial predictability, and subtraction of multiple events: *The Leading Edge*, **18**, no. 1, 55–58.

Ground roll and the Radial Trace Transform – revisited

Morgan Brown and Jon Claerbout¹

ABSTRACT

The Radial Trace Transform (RTT) is an attractive tool for wavefield separation because it lowers the apparent temporal frequency of radial events like ground roll, making it possible to remove them from the data by simple bandpass filtering in the Radial Trace (RT) domain. We discuss two implementations of the RTT. In the first, and better known, the RT domain is well-sampled, and thus suitable for post-filtering, but is prone to interpolation errors. We present an alternate implementation, which is pseudo-unitary in the limit of an infinitely densely sampled RT space, with the side effect that the RT domain has missing data. Using a simple 2-D filter for regularization, we estimate the missing data in the RT domain by least squares optimization, without affecting the invertibility of the RTT. Our implementation suppresses radial noise while preserving signal, including static shifts. Although it runs into trouble when noise is spatially aliased, we show that application of a linear moveout correction prior to processing increases our scheme's effectiveness.

INTRODUCTION

The Radial Trace Transform (RTT), is a simple coordinate transform of normal (t, x) domain seismic gathers; a horizontal deformation, accomplished by the following linear mapping

$$\begin{aligned} t &\rightarrow t \\ x &\rightarrow v = \frac{x}{t} \end{aligned} \tag{1}$$

The radial coordinate is termed “ v ” because the RTT sorts the data by apparent velocity. Neglecting dispersion effects, ground roll maps to zero temporal frequency in the RT domain. The RTT is not a new concept. Nearly twenty years ago, this coordinate transform was an active subject of research for use in multiple suppression (Taner, 1980), migration (Ottolini, 1982), and even for the subject of this paper, ground roll removal (Claerbout, 1983). Henley's (1999) recent paper reminded the world of the RTT's usefulness in attacking ground roll. Sava and Fomel (2000) use the RTT to compute angle gathers, exploiting the fact that slant-stack in the (t, x) domain is equivalent to computing the RTT in the Fourier domain.

Figure 1 shows 30 radial traces (thick lines) overlain on a rectangular (t, x) mesh. Points along radial traces rarely fall exactly on (t, x) data locations, making the RTT an exercise in

¹email: morgan@sep.stanford.edu, claerbout@stanford.edu

interpolation. In matrix notation, let us denote the RTT of data \mathbf{d} as follows:

$$\mathbf{R}\mathbf{d} = \mathbf{r} \tag{2}$$

\mathbf{R} is generally non-square, and even if square, is usually noninvertible. Often, however, such interpolation operators are nearly unitary: $\mathbf{R}^T\mathbf{R} \approx \mathbf{I}$. We define the interpolation error as follows

$$\mathbf{e} = [\mathbf{I} - \mathbf{R}^T\mathbf{R}]\mathbf{d}. \tag{3}$$

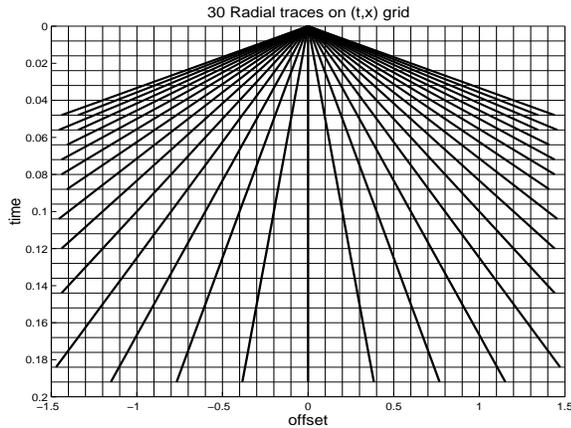


Figure 1: 30 radial traces overlaying a (t, x) grid. morgan1-figure1 [NR]

METHODOLOGY

In this paper, two implementations of the RTT are investigated. Each is illustrated schematically in Figure 2.

1. **v -interpolation method:** For fixed (t, x) bins, linearly interpolate between radial traces.
2. **x -interpolation method:** For fixed (t, v) bins, linearly interpolate between offsets.

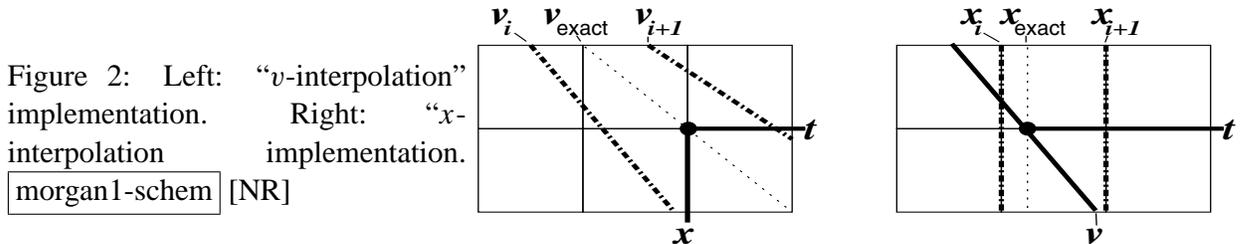


Figure 2: Left: “ v -interpolation” implementation. Right: “ x -interpolation” implementation. morgan1-schem [NR]

In the v -interpolation approach, \mathbf{R} “pushes” energy — weighted by linear interpolation coefficients — from fixed (t, x) bins into the two radial trace (t, v) bins that bracket them. This implementation will cause interpolation error in regions where many (t, x) bins lie between

adjacent radial traces (see Fig. 1). However, the RT domain is nonphysical in the sense that as many radial traces can be used as computer memory permits, so the interpolation error can essentially be driven to zero by sampling densely enough in RT space. Unfortunately, by sampling finely, in many regions of RT space there are pairs of radial traces which bracket no (t, x) bin, so “holes” are introduced into the RT space which inhibit later filtering operations.

In the x -interpolation approach, \mathbf{R} “pulls” energy into fixed (t, v) bins from the two offset (t, x) bins that bracket them. The interpolation error of this implementation depends only on the trace spacing of the data. The net effect of applying the operator is to smooth laterally, making this implementation dangerous if the data has even small static shifts. At typical trace spacing, \mathbf{R} is not pseudo-unitary, but since the RT space is guaranteed to have no “holes”, this implementation is appropriate for post-filtering operations. Henley (1999) used the x -interpolation approach, as did the author of the SEPlib program `Radial`.

The two implementations of the RTT discussed above illustrate a fundamental, and oft-ignored duality in the analysis of interpolation operators. Intuition supports the x -interpolation method — any given unknown model point is the weighted average of the two known data points which bracket it. In applications like NMO, where averaging is done along the well-sampled time axis, this intuition is sensible, but it breaks down when the averaging is across offsets. The alternate approach (v -interpolation) is less intuitive, as the interpolation is done across radial traces, in the “virtual” space of the model. Since the model space is not constrained by the parameters of data acquisition, it can be sampled as densely as needed to minimize interpolation error. In fact, in the limit of infinitely dense sampling in model space, simple nearest neighbor binning drives interpolation error to zero.

The idea of this paper is to use the v -interpolation RTT to take advantage of its minimal interpolation error, and then handle the problem of “holes” in RT space by missing data estimation. Ideally, the seismic gather contains nearly-horizontal primary reflections and radial events, so the RT space is composed of nearly-vertical $(h(v))$ and nearly-horizontal $(l(t))$ events which a cascade of derivative operators extinguishes:

$$\frac{\partial}{\partial v} \frac{\partial}{\partial t} [h(v) + l(t)] = 0 \quad (4)$$

A finite difference stencil approximating $\partial^2/\partial v \partial t$ is

$$\mathbf{a} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}. \quad (5)$$

Filter \mathbf{a} is well-suited for the helical least-squares missing data estimation methodology of Claerbout (1999). The missing data problem is driven by two fitting goals: 1) Honor the known data points exactly, and 2) impose any prior knowledge on the unknown model parameters via regularization. Define a known data mask \mathbf{K} , simply a diagonal operator of the same dimension as \mathbf{m} ; 1 where data is known, 0 otherwise. The prior assumptions about the model are contained in filter \mathbf{a} . Define \mathbf{A} as the convolution matrix which applies the filter \mathbf{a} and combine the two fitting goals into a single regularized optimization problem.

$$\mathbf{K}\mathbf{m} = \mathbf{K}\mathbf{R}\mathbf{d} \quad (6)$$

$$\epsilon \mathbf{A}\mathbf{m} \approx 0. \quad (7)$$

Equation (6) forces the model to match the data where the latter is known. Equation (7) minimizes the power of the convolution of \mathbf{a} with \mathbf{m} , i.e., optimality is achieved when the unknown model is filled with horizontal and vertical events. ϵ is a user-chosen scale factor.

In this paper, we seek to use the RTT to do noise suppression. Specifically, we map a 2-D seismic gather to RT space, then apply a conservative (6.5 Hz cutoff) highpass filter – call it \mathbf{B} – to remove the noise, and finally transform back to (t, x) space. In symbols, we can write the estimated signal, $\hat{\mathbf{s}}$, as follows

$$\hat{\mathbf{s}} = \mathbf{R}^T \mathbf{B} \mathbf{R} \mathbf{d}. \quad (8)$$

The corresponding estimate of the noise, $\hat{\mathbf{n}}$, is obtained simply by subtracting the estimated signal from the data:

$$\hat{\mathbf{n}} = \mathbf{d} - \hat{\mathbf{s}} = (\mathbf{I} - \mathbf{R}^T \mathbf{B} \mathbf{R}) \mathbf{d}. \quad (9)$$

RESULTS

Figure 3 shows a 2-D shot gather from a multicomponent survey in Venezuela, which exhibits dispersive ground roll with an apparent velocity range of 200-500 m/s. A relatively fine trace spacing of 17 m partially mitigates spatial aliasing. Some weak backscattered energy may be present. Additionally, static shifts of \pm one sample were introduced to the data randomly, in order to emphasize the effect of the RTT on data whose lateral coherency may be degraded. Figure 4 shows the RTT of the data in Figure 3 for both implementations. Each panel contains 300 radial traces. As mentioned by Henley (1999), the origin of the RTT can be placed anywhere. Often in land surveys, the near-zero offset traces are not recorded, meaning that the apparent origin of radial events will lie off the section. The following results were obtained with the the origin of the RTT placed at 0.25 seconds. As expected, the raw v -interpolation panel has many “holes”, while the x -interpolation panel does not. The missing data estimation algorithm described above has plausibly filled the missing data, insofar as visual similarity to the x -interpolation panel is a valid measure.

Figure 5 compares the error arising from both implementations of the RTT (equation (3)). In the RTT panels shown in Figure 4, each containing 300 radial traces, the error in both of the v -interpolation panels is negligible. This figure gives visual proof of the fact that the missing data infill process does not harm the original data, i.e., that the missing data points in RT space are in the nullspace of \mathbf{R}^T . As expected, the x -interpolation error is nonzero, particularly for high-wavenumber events like ground roll and likely backscatter. Additionally, and less obviously, the primary events between 1.75 and 2.5 seconds suffer energy loss and a noticeable lateral smoothing as a result of the transform.

Figure 6 is the estimated signal (equation (8)). Notice that missing data infill has markedly improved the quality of noise suppression obtained by the v -interpolation technique. Still, by visual inspection, we must conclude that the x -interpolation result is the best of the three for noise suppression.

Figure 3: 2-D shot gather.
morgan1-hector-dat [ER]

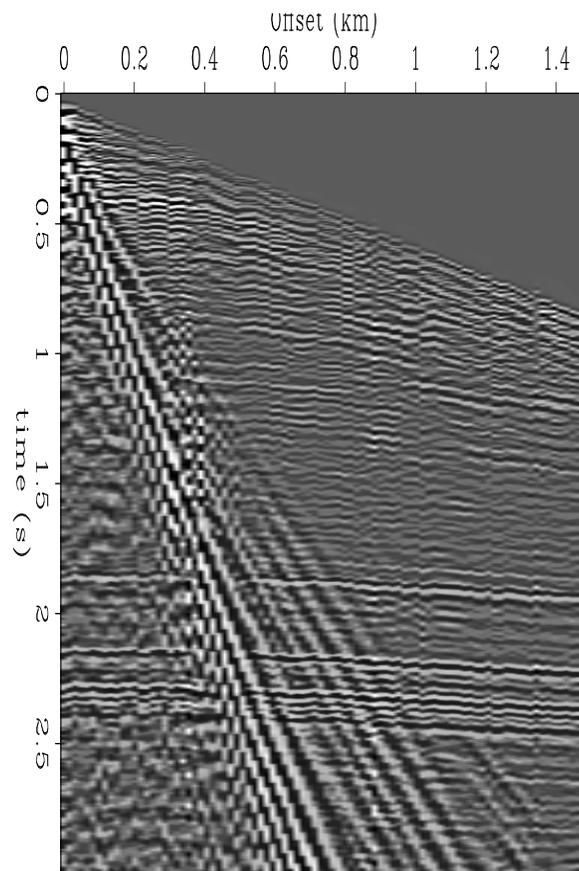


Figure 7 is the estimated noise (equation (9)). First notice that x -interpolation result contains more than just the radial ground roll that the simple physical model accounts for. From Figure 5, we know that the presence of the primary energy between 1.75 and 2.5 seconds and backscattered noise in the x -interpolation panel is due to interpolation errors, and not to the highpass filtering. The non-infilled v -interpolation result contains some energy (either artifacts or primary energy) around 1 second, while the infilled result does not. From the standpoint of signal preservation, the v -interpolation result with missing data infill is the best of the three. Philosophically, by using a pseudo-unitary RTT operator and thus ensuring that the only thing modifying the original data is the bandpass filter, the v -interpolation implementation honors the physics which drives the problem in the first place.

Aliased Data

Ground roll is nearly always spatially aliased, so the relatively unaliased example of Figure 3 is a somewhat unrealistic exception to the practical rule. To inject some realism, we decimated the original 2-D shot gather (Figure 3) by a factor of two in offset, as shown in Figure 8, so that the ground roll is quite aliased. Figure 9 compares the RTT of the decimated data. The results are disappointing. Looking at the v -interpolation without infill panel (top), the human eye can easily interpolate vertically to reconstruct the radial events in RT space. Unfortunately, the v -interpolation panel with infill does not have the desired vertical coherence. In fact, it would seem that the central premise motivating this paper — that the RTT maps ground roll to zero temporal frequency — is violated. Figures 10 and 11 are analogous to Figures 6 and 7 — they are the estimates of signal and noise, respectively. All implementations (v -interpolation with and without infill, and x -interpolation) do an relatively poor job of noise suppression.

A simple way to dealias linear ground roll is to apply a linear moveout (LMO) correction. Figure 12 shows the result of applying a 1.5 km/sec LMO correction to the decimated data of Figure 8. The ground roll is no longer spatially aliased, but the primaries are also no longer “flat”, as they were originally. As a result, interpolation errors for the x -interpolation RTT will increase. Figure 13 compares the RTT panels for the decimated/LMO’ed data. The ground roll now occupies a higher effective velocity band, and more importantly, is much closer to zero temporal frequency than in Figure 9. The noise suppression achieved (Figure 14) is better than the case in which LMO was not used (Figure 10). As expected, and mentioned above, the x -interpolation RTT leads to severe losses of signal energy, quite a bit more severe than either of the two v -interpolation implementations, as can be seen in Figure 15. Unfortunately, both v -interpolation implementations seem to suffer some small signal losses, which suggests that LMO may actually be “aliasing” the primaries by mapping them to low temporal frequency in the RT domain.

CONCLUSIONS

Our implementation of the RTT effectively suppresses unaliased radial noise while preserving signal, including static shifts. In its current form, our missing data interpolation technique did

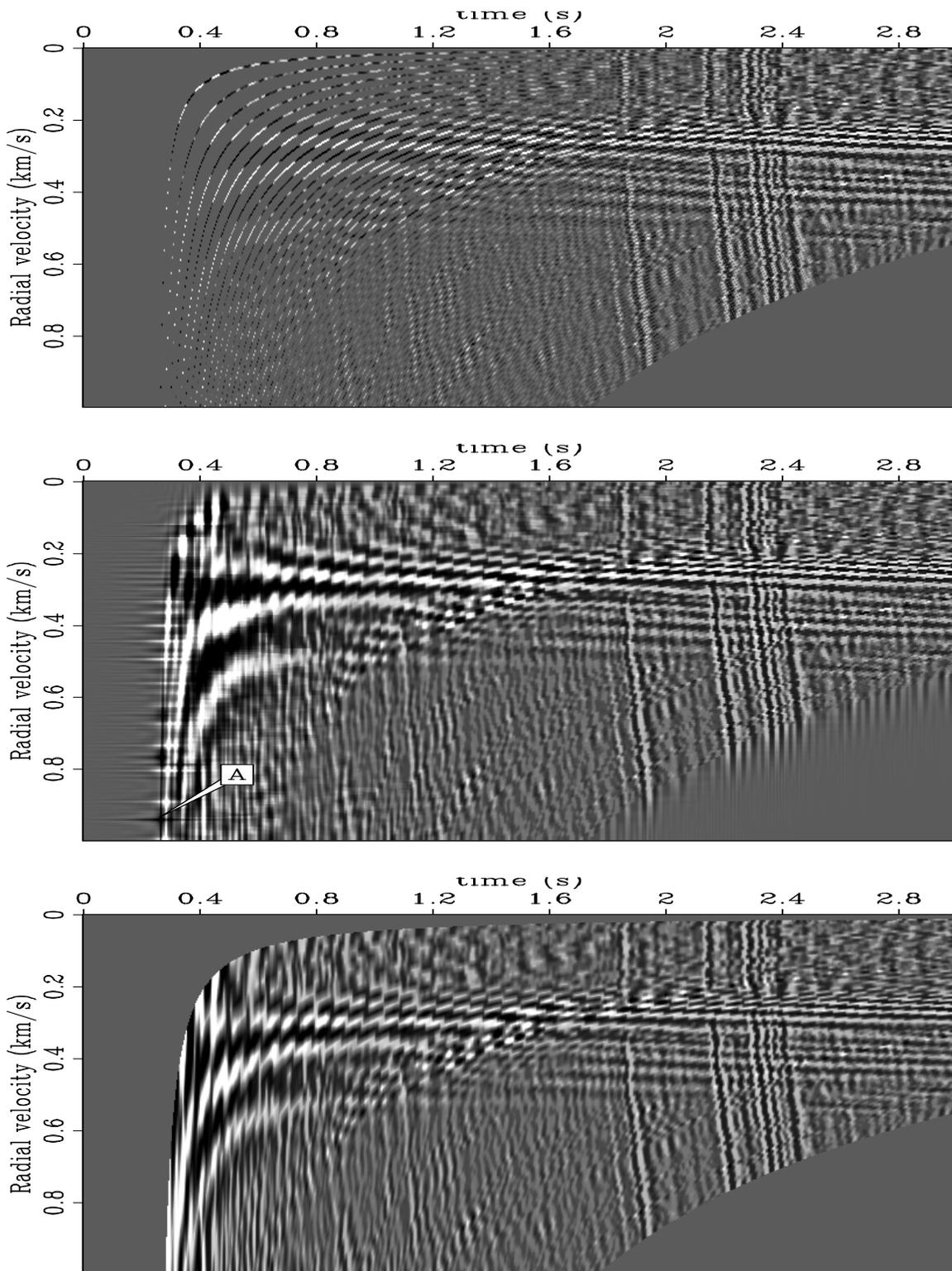


Figure 4: Radial Trace Transform. Top: v -interpolation without infill. Middle: v -interpolation with infill. "A" points to an example of the "+"-shaped impulse response of the missing data filter of equation (7). Bottom: x -interpolation. [morgan1-hector-radial-comp](#) [ER,M]

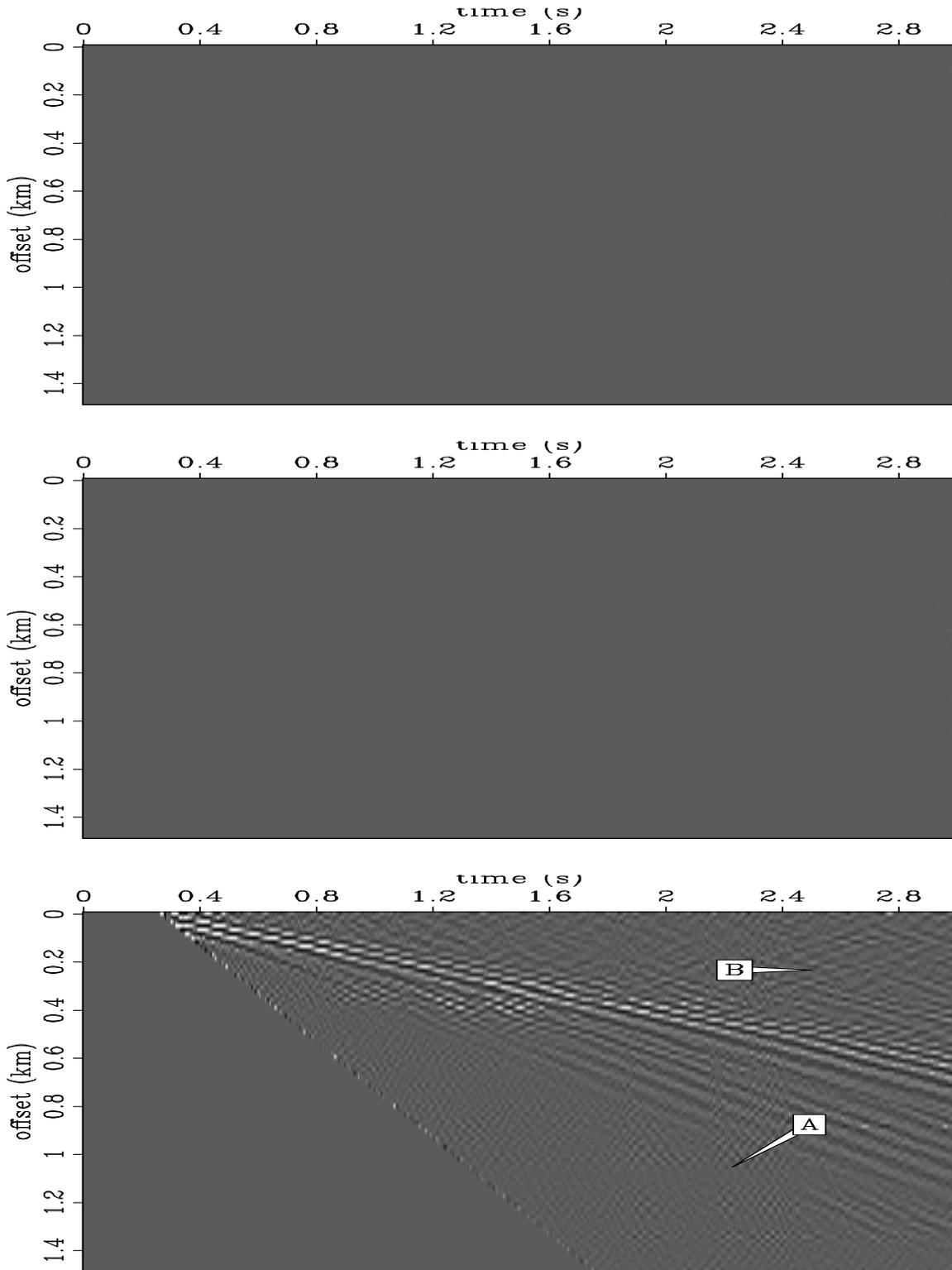


Figure 5: Interpolation error. Top: v -interpolation error without infill. Middle: v -interpolation error with infill. Bottom: x -interpolation error. “A” points to lost energy from the primary events around 2 seconds. “B” points to removed backscattered noise.

morgan1-hector-raderr-comp [ER,M]

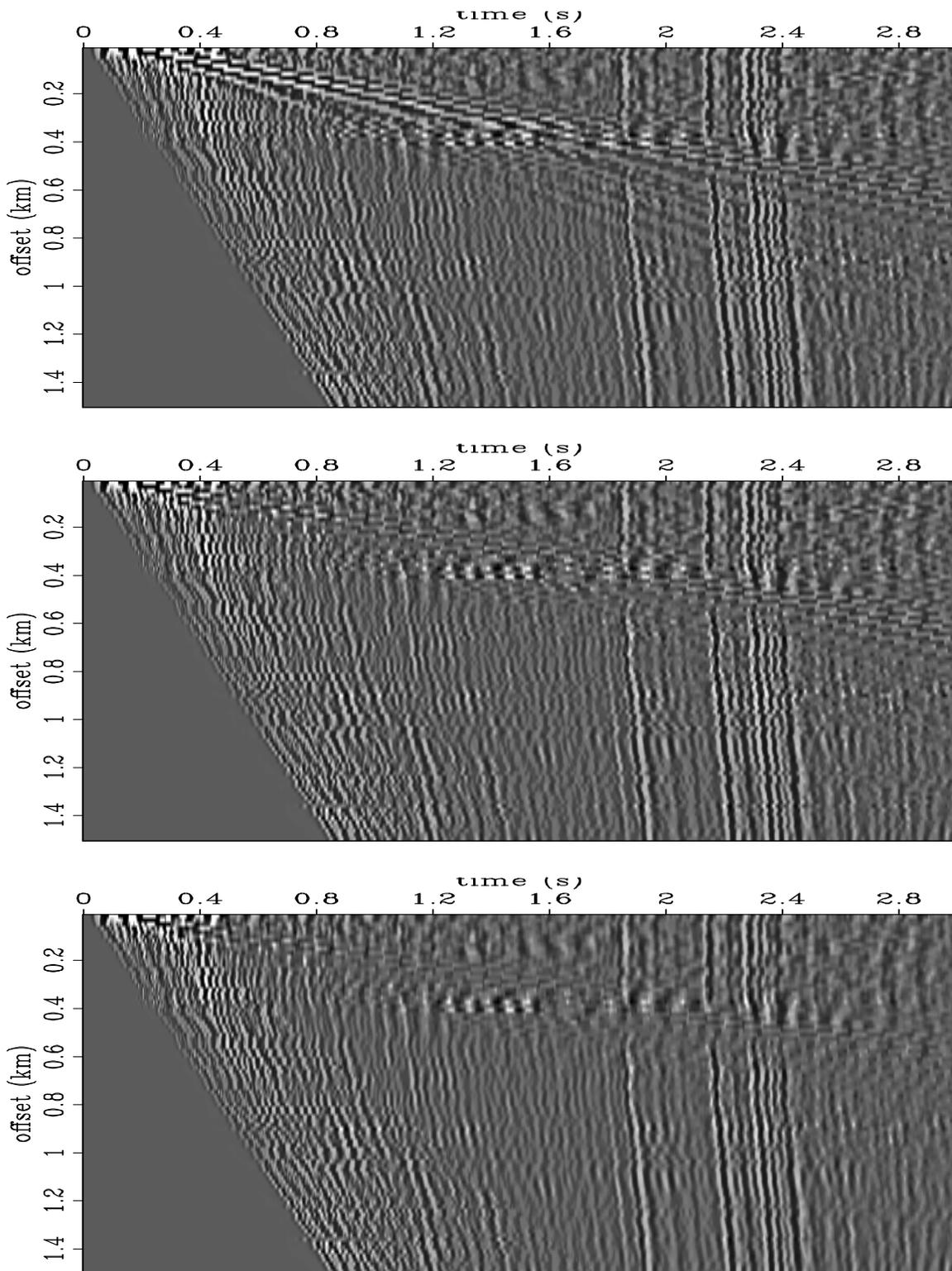


Figure 6: Estimated signal. Top: v -interpolation without infill. Middle: v -interpolation with infill. Bottom: x -interpolation. [morgan1-hector-estsig](#) [ER,M]

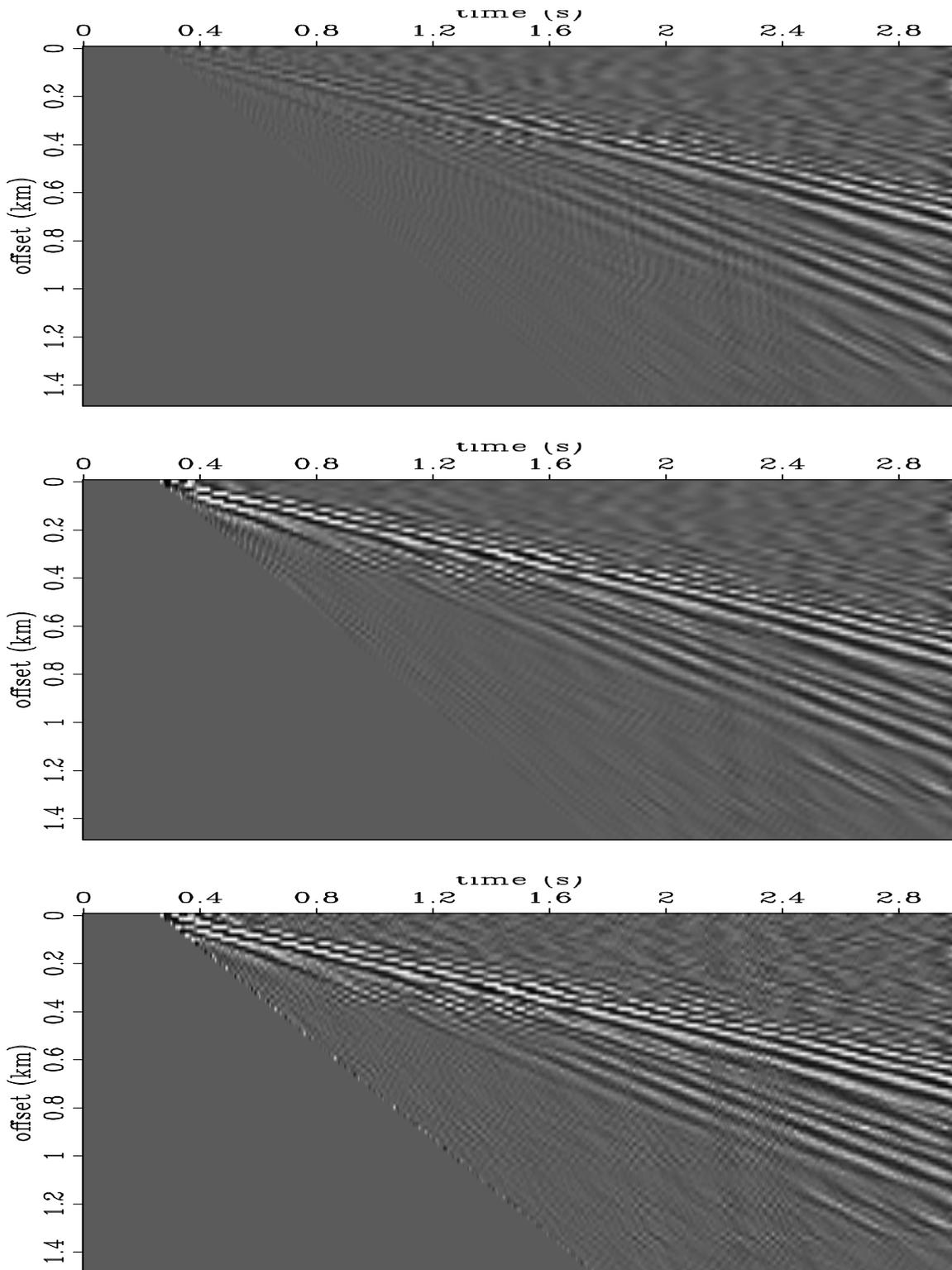
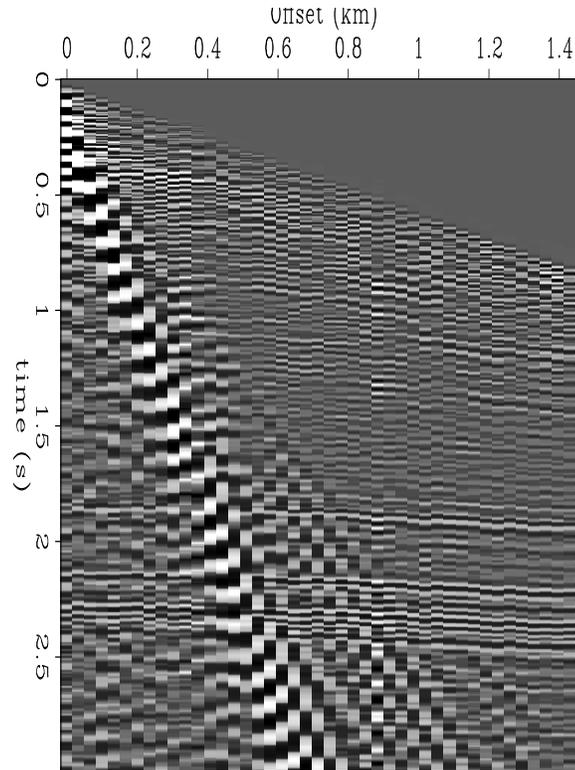


Figure 7: Estimated noise. Panels defined as in Figure 6. `morgan1-hector-estnoiz` [ER,M]

Figure 8: Same 2-D shot gather as Figure 3, only decimated by a factor of two in offset.

morgan1-hectoralias-dat [ER]



a relatively poor job of coherently interpolating spatially aliased radial noise events to vertical events in the RT domain, although we have high hopes that success is only a small conceptual leap away. To combat aliasing, we applied an LMO correction to the data to dealias the noise events with an LMO correction, leading to improved noise suppression, at the cost of some lost signal.

REFERENCES

- Claerbout, J. F., 1983, Ground roll and radial traces: SEP-35, 43–54.
- Claerbout, J., 1999, Geophysical estimation by example: Environmental soundings image enhancement: Stanford Exploration Project, <http://sepwww.stanford.edu/sep/prof/>.
- Henley, D., 1999, The radial trace transform: an effective domain for coherent noise attenuation and wavefield separation: 69th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, pages 1204–1207.
- Ottolini, R., 1982, Migration of reflection seismic data in angle midpoint coordinates: SEP-33.
- Sava, P., and Fomel, S., 2000, Angle-gathers by Fourier Transform: SEP-103, 119–130.
- Taner, M. T., 1980, Long-period sea-floor multiples and their suppression: Geophys. Prosp., 28, no. 01, 30–48.

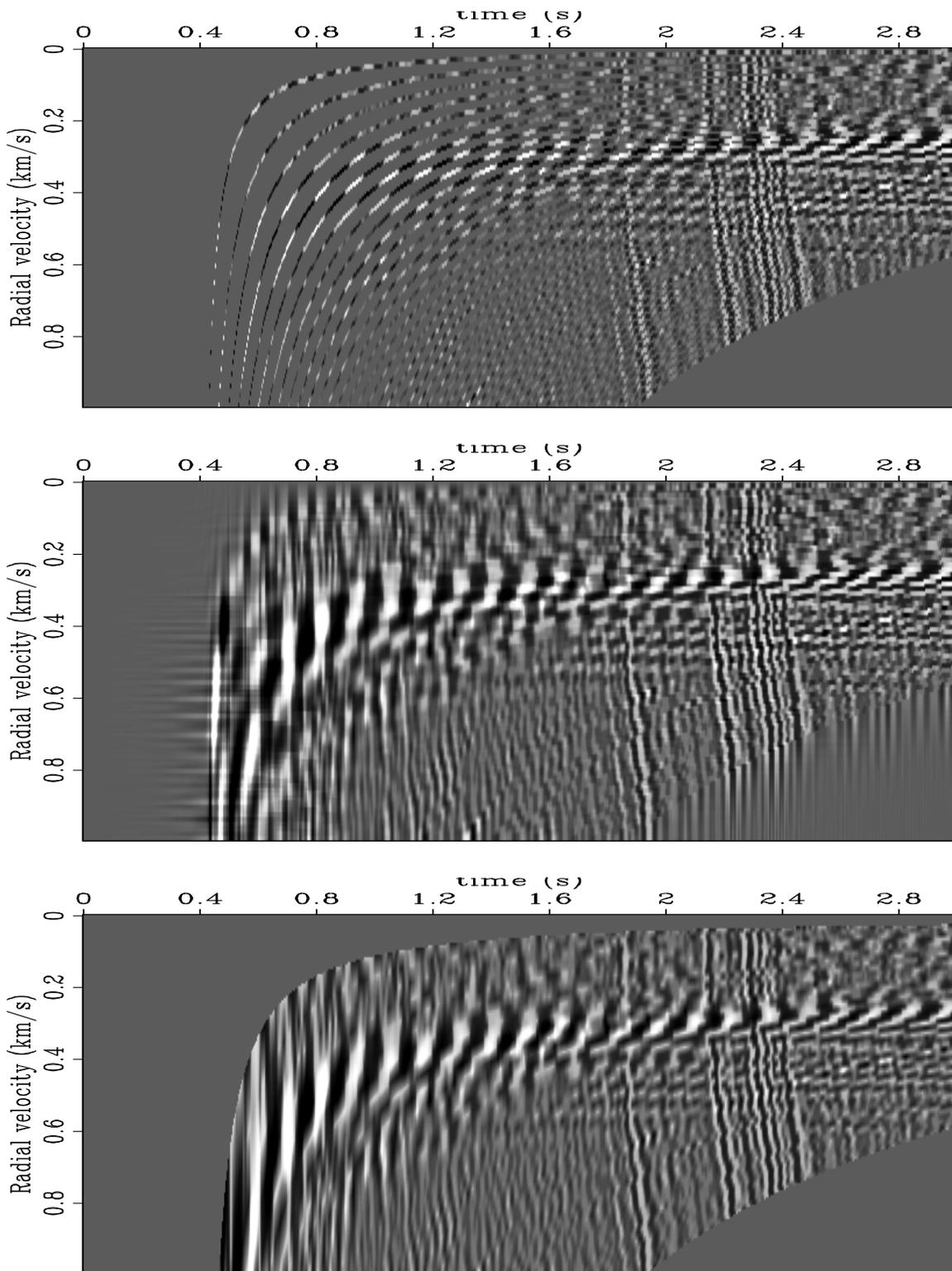


Figure 9: Top: v -interpolation without infill. Middle: v -interpolation with infill. Bottom: x -interpolation. `morgan1-hectoralias-radial-comp` [ER,M]

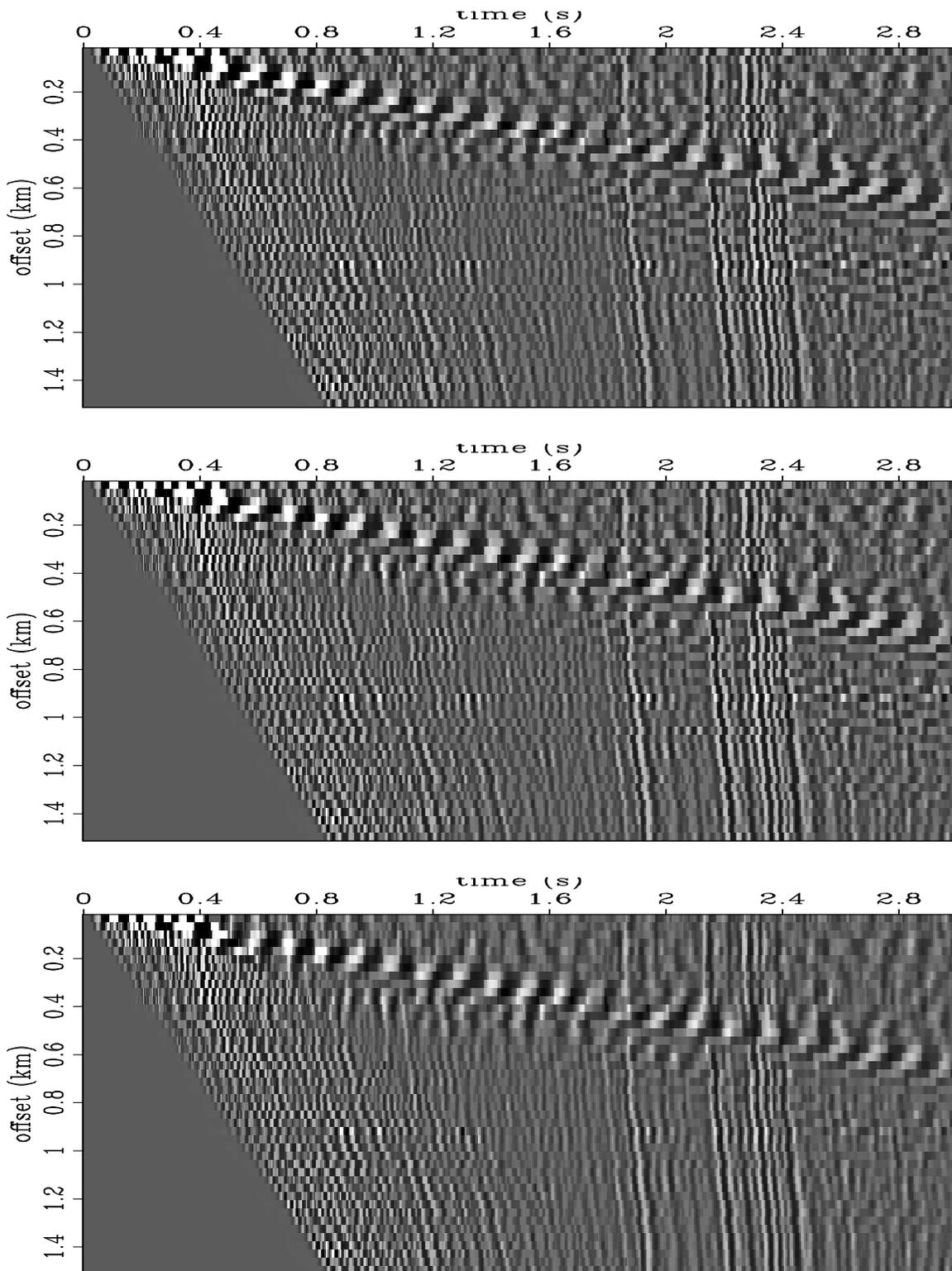


Figure 10: Estimated signal. Top: v -interpolation without infill. Middle: v -interpolation with infill. Bottom: x -interpolation. [morgan1-hectoralias-estsig](#) [ER,M]

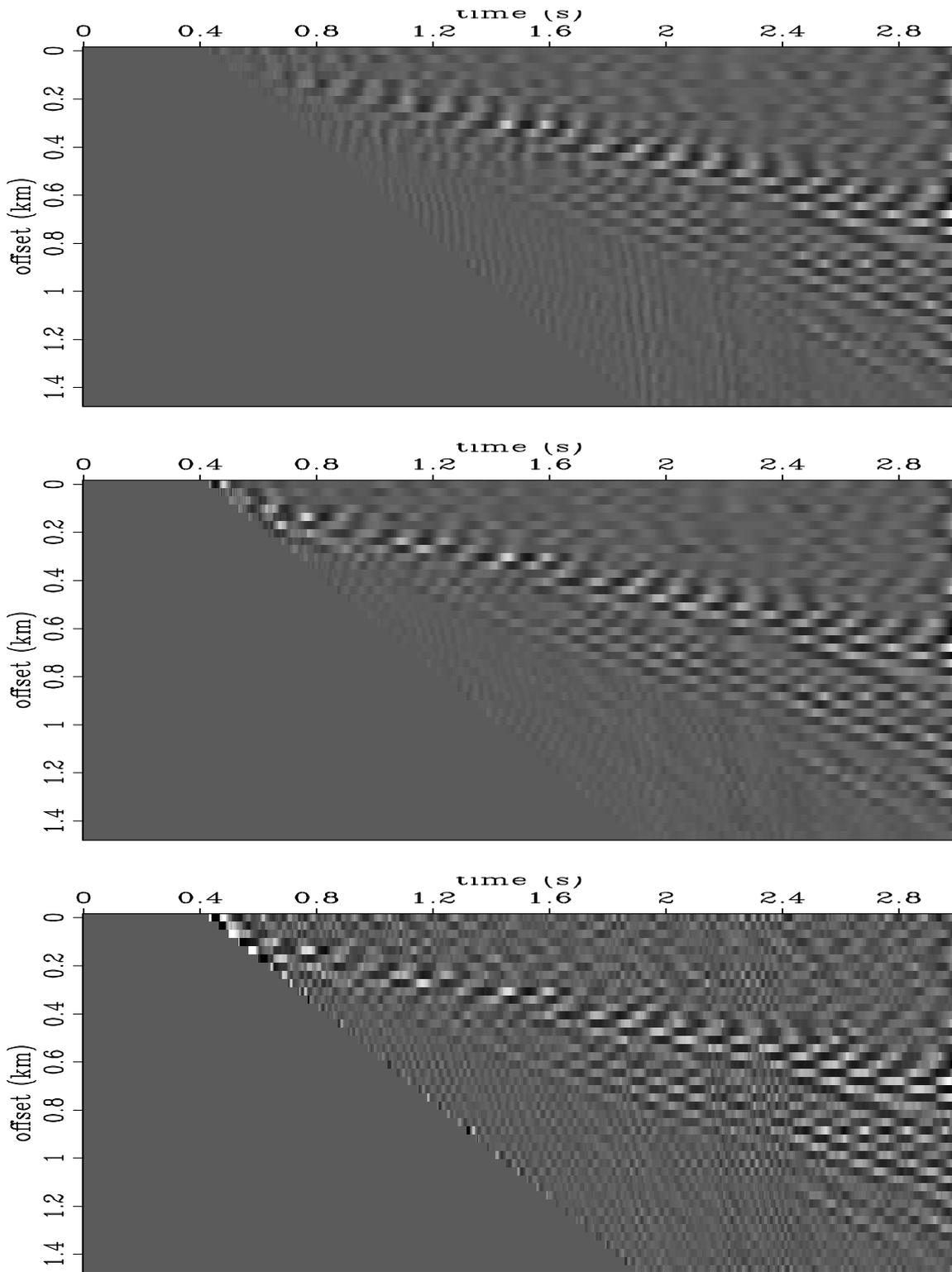
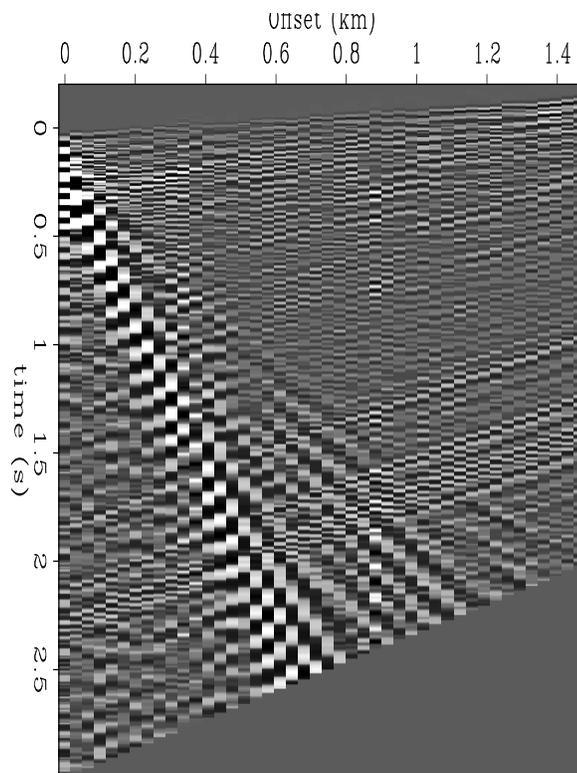


Figure 11: Estimated noise. Panels defined as in Figure 10.
[ER,M]

morgan1-hectoralias-estnoiz

Figure 12: Decimated 2-D shot gather (Figure 8), after 1.0 km/sec linear moveout correction. `morgan1-hectorlmo-dat`
[ER]



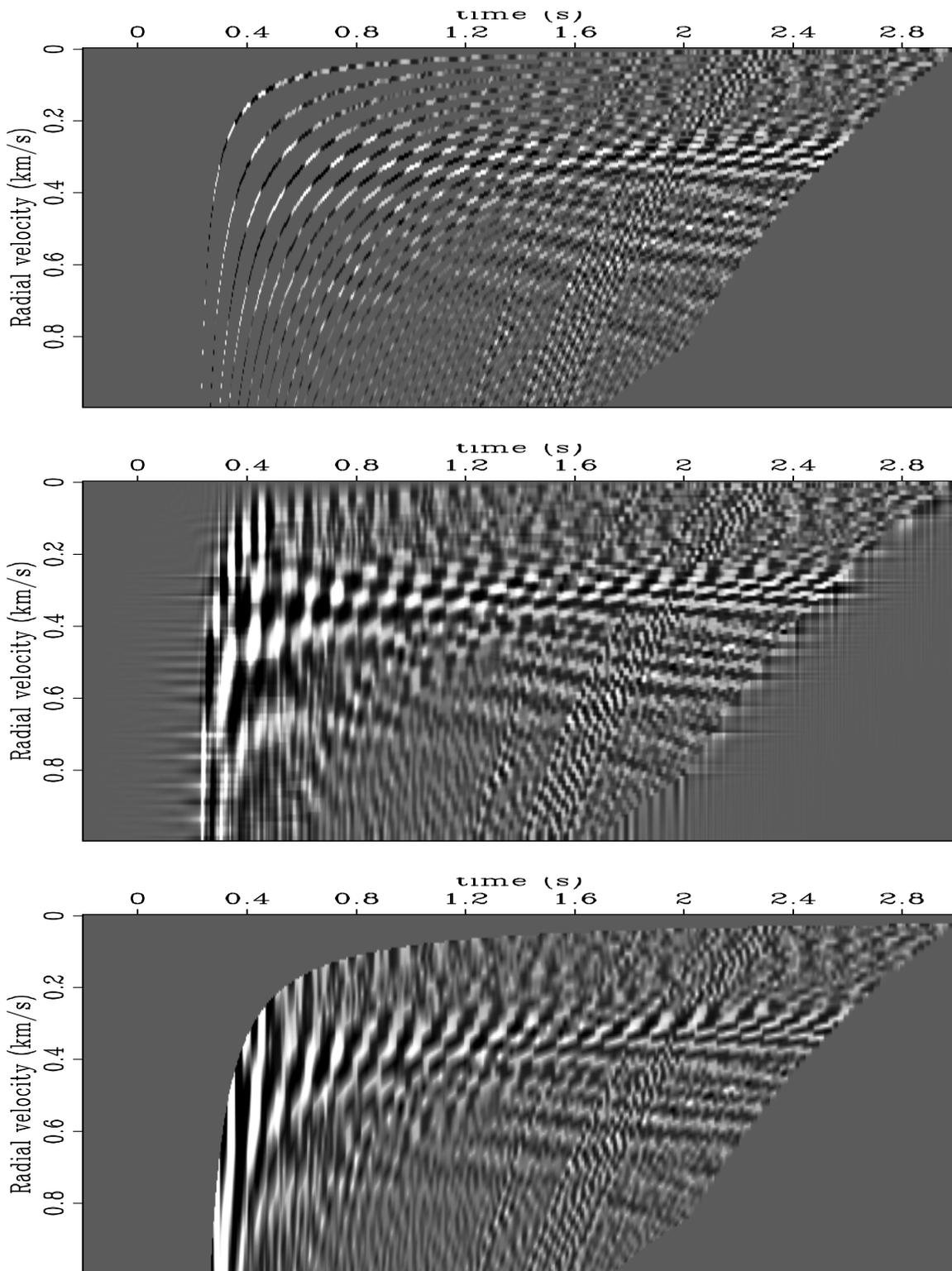


Figure 13: Top: v -interpolation without infill. Middle: v -interpolation with infill. Bottom: x -interpolation. `morgan1-hectorlmo-radial-comp` [ER,M]

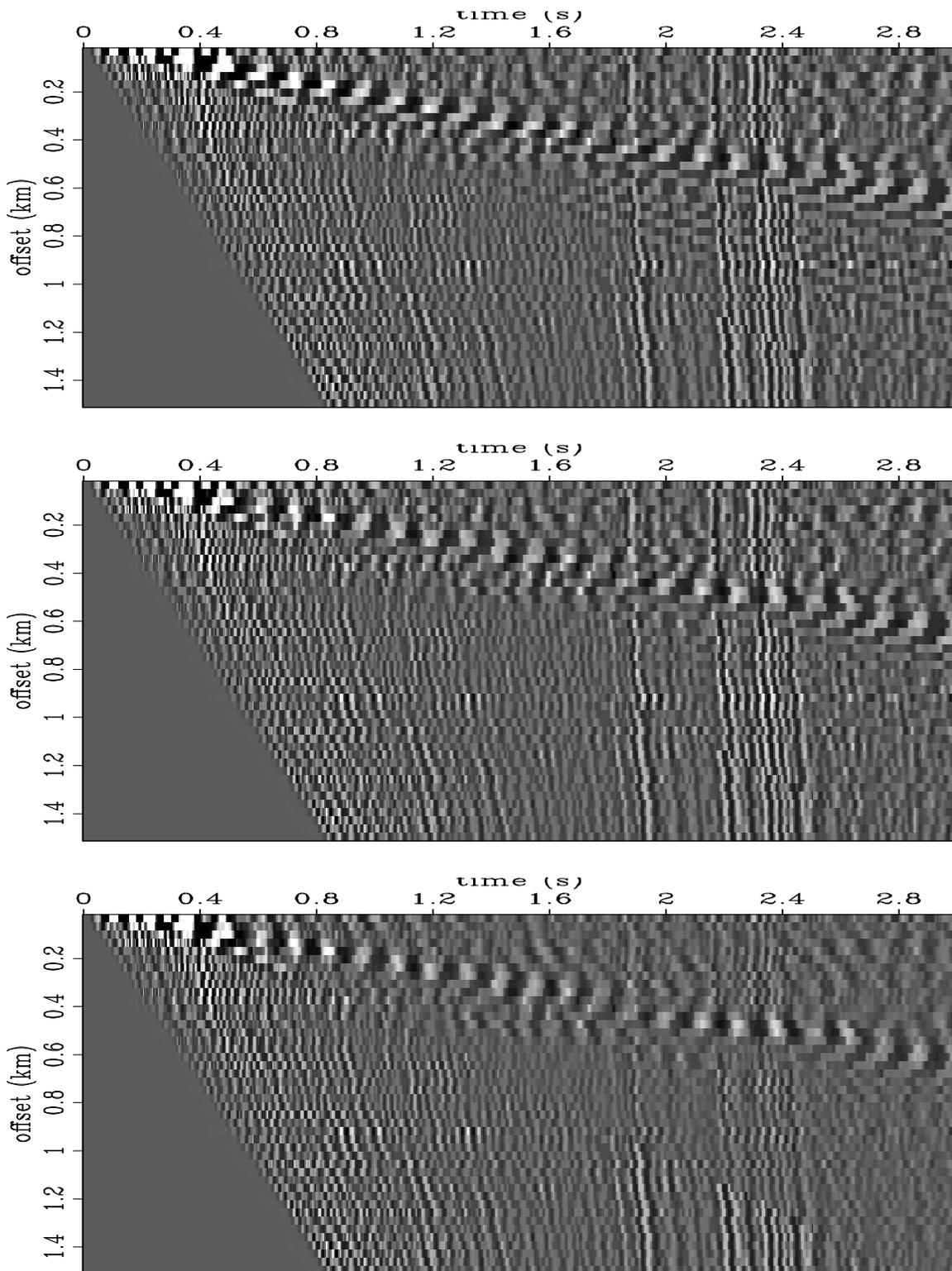


Figure 14: Estimated signal. Top: v -interpolation without infill. Middle: v -interpolation with infill. Bottom: x -interpolation. [morgan1-hectorlmo-lmo-estsig](#) [ER,M]

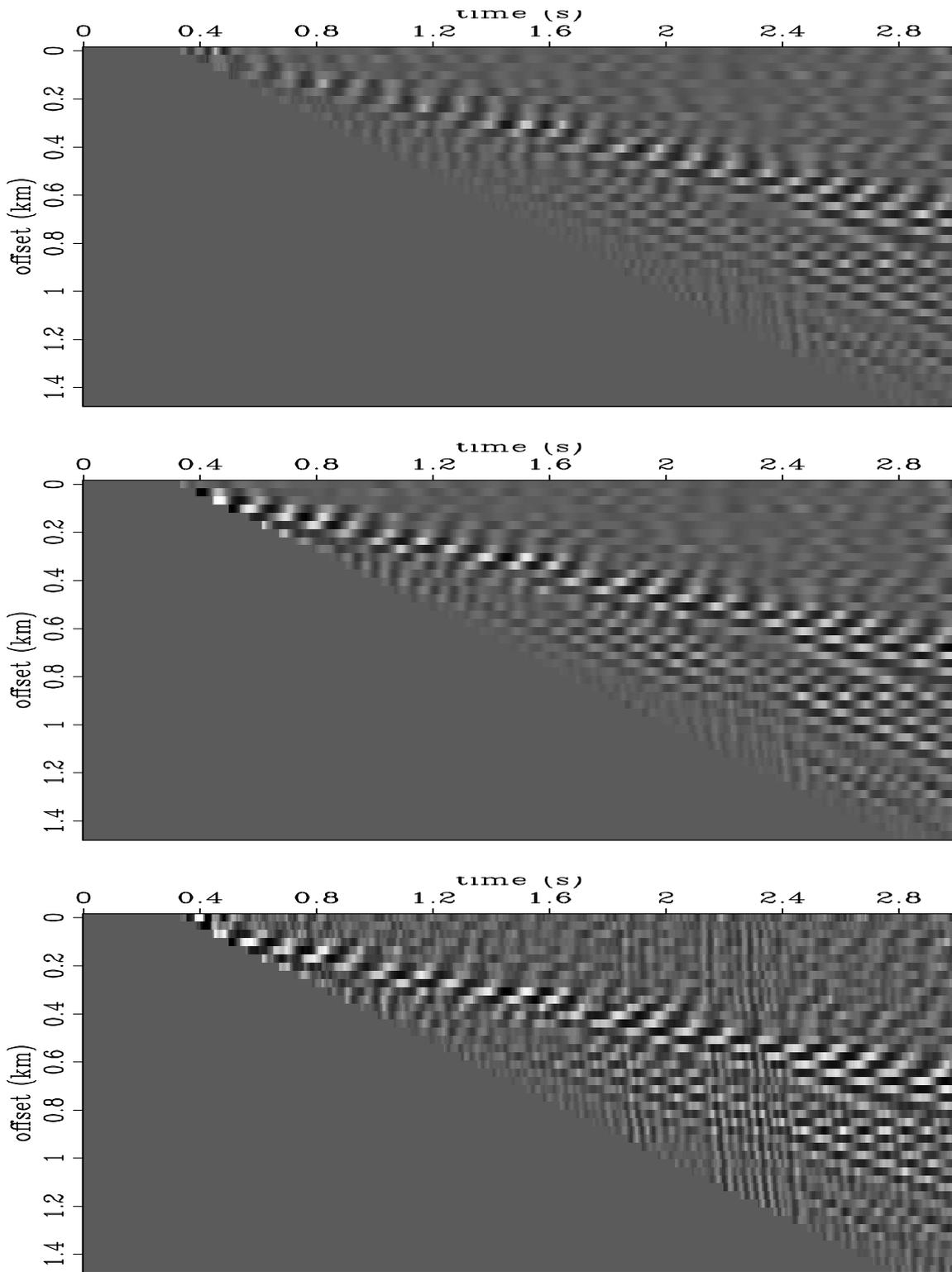


Figure 15: Estimated noise. Panels defined as in Figure 14.

morgan1-hectorlmo-lmo-estnoiz

[ER,M]

Seismic data interpolation with the offset continuation equation

*Sergey Fomel*¹

ABSTRACT

I propose a finite-difference offset continuation filter for interpolating seismic reflection data. The filter is constructed from the offset continuation differential equation and is applied on frequency slices in the log-stretch frequency domain. Synthetic data tests produce encouraging results: nearly perfect interpolation of a constant-velocity dataset with a complex reflector model and reasonably good interpolation of the Marmousi dataset.

INTRODUCTION

As early as 20 years ago, researchers at SEP considered data interpolation as one of the most important problems of seismic data processing (Claerbout, 1980, 1981; Thorson, 1981). In 2-D exploration, the interpolation problem arises because of missing near and far offsets, spatial aliasing and occasional bad traces. In 3-D exploration, the importance of this problem increases dramatically because 3-D acquisition almost never provides a complete regular coverage in both midpoint and offset coordinates (Biondi, 1999). Data regularization in 3-D can solve the problem of Kirchhoff migration artifacts (Gardner and Canning, 1994), prepare the data for common-azimuth imaging (Biondi and Palacharla, 1996), or provide the spatial coverage required for 3-D multiple elimination (van Dedem and Verschuur, 1998).

Claerbout (1992, 1999) formulates the following general principle of missing data interpolation:

A method for restoring missing data is to ensure that the restored data, after specified filtering, has minimum energy.

How can one specify an appropriate filtering for a given interpolation problem? Smooth surfaces are conveniently interpolated with Laplacian filtering (Briggs, 1974; Fomel, 2000). Steering filters help us interpolate data with predefined dip fields (Clapp et al., 1997; Fomel et al., 1997; Fomel, 1999). Prediction-error filters in time-space or frequency-space domain successfully interpolate data composed of distinctive plane waves (Spitz, 1991; Claerbout, 1999). Because prestack seismic data is not stationary in the offset direction, non-stationary prediction-error filters need to be estimated, which leads to an accurate but relatively expensive method with many adjustable parameters (Crawley, 1999; Clapp et al., 1999).

¹email: sergey@sep.stanford.edu

A simple model for reflection seismic data is a set of hyperbolic events on a common midpoint gather. The simplest filter for this model is the first derivative in the offset direction applied after the normal moveout correction.² Going one step beyond this simple approximation requires taking the dip moveout (DMO) effect into account (Deregowski, 1986). The DMO effect is fully incorporated in the offset continuation differential equation (Fomel, 1994, 1995a).³

Offset continuation is a process of seismic data transformation between different offsets (Deregowski and Rocca, 1981; Bolondi et al., 1982; Salvador and Savelli, 1982). Different types of DMO operators (Hale, 1995) can be regarded as continuation to zero offset and derived as solutions of an initial-value problem with the revised offset continuation equation (Fomel, 1995b). Within a constant-velocity assumption, this equation not only provides correct traveltimes on the continued sections, but also correctly transforms the corresponding wave amplitudes (Fomel and Bleistein, 1996). Integral offset continuation operators have been derived independently by Stovas and Fomel (1996), Bagaini and Spagnolini (1996), and Chemingui and Biondi (1994). The 3-D analog is known as azimuth moveout (AMO) (Biondi et al., 1998). In the shot-record domain, integral offset continuation transforms to shot continuation (Schwab, 1993; Bagaini and Spagnolini, 1993; Spagnolini and Opreni, 1996). Integral continuation operators can be applied directly for missing data interpolation and regularization (Bagaini et al., 1994; Mazzucchelli and Rocca, 1999). However, they don't behave well for continuation at small distances in the offset space because of limited integration apertures and, therefore, are not well suited for interpolating neighboring records. Additionally, as all integral (Kirchoff-type) operators they suffer from irregularities in the input geometry. The latter problem is addressed by accurate but expensive inversion to common offset (Chemingui, 1999).

In this paper, I propose an application of offset continuation in the form of a finite-difference filter for Claerbout's method of missing data interpolation. The filter is designed in the log-stretch frequency domain, where each frequency slice can be interpolated independently. Small filter size and easy parallelization among different frequencies assure a high efficiency of the proposed approach. Although the offset continuation filter lacks the predictive power of non-stationary prediction-error filters, it is much simpler to handle and serves as a good *a priori* guess of an interpolative filter for seismic reflection data. I test the proposed method by interpolating randomly missing traces in a constant-velocity synthetic and by restoring near offsets and intermediate shot gathers in the Marmousi synthetic dataset. These early tests produce encouraging results. In the final section of the paper, I discuss possible strategies for improving the method.

²A similar filter appears in velocity estimation with the differential semblance method (Symes and Carazzone, 1991; Symes, 1999).

³To the author's knowledge, the first derivation of the revised offset continuation equation was done by Joseph Higginbotham of Texaco in 1989. Unfortunately, Higginbotham's derivation never appeared in open literature.

PROBLEM FORMULATION

If \mathbf{D} is a regularization operator, and \mathbf{m} is the estimated model, then Claerbout's interpolation method amounts to minimizing the power of \mathbf{Dm} ($\mathbf{m}^T \mathbf{D}^T \mathbf{Dm}$) under the constraint

$$\mathbf{K}\mathbf{m} = \mathbf{m}_k, \quad (1)$$

where \mathbf{m}_k stands for the known data values, and \mathbf{K} is a diagonal matrix with 1s at the known data locations and zeros elsewhere. It is easy to implement a constraint of the form (1) in an iterative conjugate-gradient scheme by simply disallowing the iterative process to update model parameters at the known data locations (Claerbout, 1999).

The operator \mathbf{D} can be considered as a differential equation that we assume the model to satisfy. If \mathbf{D} is able to remove all correlated components from the model and produce white Gaussian noise in the output, then $\mathbf{D}^T \mathbf{D}$ is essentially equivalent to the inverse covariance matrix of the model, which appears in the statistical formulation of least-squares estimation (Tarantola, 1987).

In this paper, I propose to use the offset continuation equation (Fomel, 1995a) for the operator \mathbf{D} . Under certain assumptions, this equation is indeed the one that prestack seismic reflection data can be presumed to satisfy. The equation has the following form:

$$h \left(\frac{\partial^2 P}{\partial y^2} - \frac{\partial^2 P}{\partial h^2} \right) = t_n \frac{\partial^2 P}{\partial t_n \partial h}, \quad (2)$$

where $P(t_n, h, x)$ is the prestack seismic data after the normal moveout correction (NMO), t_n stands for the time coordinate after NMO, h is the half-offset, and y is the midpoint. Offset continuation has the following properties:

- Equation (2) describes an artificial process of prestack data transformation in the offset direction. It belongs to the class of linear hyperbolic equations. Therefore, the described process is a wave-type process. Half-offset h serves as a continuation variable (analogous to time in the wave equation).
- Under a constant-velocity assumption, equation (2) provides correct reflection travel-times and amplitudes at the continued sections. The amplitudes are correct in the sense that the geometrical spreading effects are properly transformed independently from the shape of the reflector. This fact has been confirmed both by the ray method approach (Fomel, 1995a) and by the Kirchhoff modeling approach (Fomel and Bleistein, 1996; Fomel et al., 1996).
- Dip moveout (DMO) (Hale, 1995) can be regarded as a particular case of offset continuation to zero offset (Deregowski and Rocca, 1981). As shown in my earlier paper (Fomel, 1995b), different known forms of DMO operators can be obtained as solutions of a special initial-value problem on equation (2).
- To describe offset continuation for 3-D data, we need a pair of equations such as (2), acting in two orthogonal projections. This fact follows from the analysis of the azimuth moveout operator (Fomel and Biondi, 1995; Biondi et al., 1998).

- A particularly efficient implementation of offset continuation results from a log-stretch transform of the time coordinate (Bolondi et al., 1982), followed by a Fourier transform of the stretched time axis. After these transforms, equation (2) takes the form

$$h \left(\frac{\partial^2 \tilde{P}}{\partial y^2} - \frac{\partial^2 \tilde{P}}{\partial h^2} \right) - i \Omega \frac{\partial \tilde{P}}{\partial h} = 0, \quad (3)$$

where Ω is the corresponding frequency, and $\tilde{P}(\Omega, h, x)$ is the transformed data (Fomel, 1995b). As in other F - X methods, equation (3) can be applied independently and in parallel on different frequency slices.

I propose to adopt a finite-difference form of the differential operator (3) for the regularization operator \mathbf{D} . A simple analysis of equation (3) shows that at small frequencies, the operator is dominated by the first term. The form $\frac{\partial^2 P}{\partial y^2} - \frac{\partial^2 P}{\partial h^2}$ is equivalent to the second mixed derivative in the source and receiver coordinates. Therefore, at low frequencies, the offset waves propagate in the source and receiver directions. At high frequencies, the second term in (3) becomes dominating, and the entire method becomes equivalent to the trivial linear interpolation in offset. The interpolation pattern is more complicated at intermediate frequencies.

TESTS

I started numerical testing of the proposed technique first on the constant velocity synthetic, where all the assumptions behind the offset continuation equation are valid. Encouraged by the results, I proceeded to tests on the Marmousi synthetic dataset, which is associated with a highly inhomogeneous velocity model.

Constant-velocity synthetic

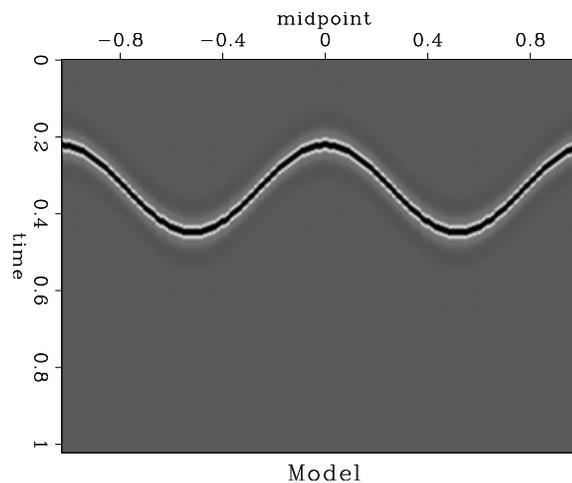


Figure 1: Reflector model for the constant-velocity test `sergey2-cup` [ER]

A sinusoidal reflector shown in Figure 1 creates complicated reflection data, shown in Figure 2 before and after the normal-moveout correction. The syncline parts of the reflector

lead to traveltimes triplications at sufficiently large offset. A mixture of different dips from the triplications would make it extremely difficult to interpolate the data in individual common-offset gathers, such as those shown in Figure 2. The plots of time slices after NMO (Figure 3) clearly show that the data are also non-stationary in the offset direction. Therefore, a simple offset interpolation scheme would also fail.

To set up an interpolation experiment, I randomly removed half of the traces in the original data and attempted to reconstruct them. Figure 4 shows the reconstruction process on individual frequency slices. Despite the complex and non-stationary character of the reflection events in the frequency domain, the offset continuation equation is able to reconstruct them quite accurately from the decimated data.

Figure 5 shows the input and the result of interpolation after transforming it back to the time domain. A comparison of the interpolation result with the ideal output (Figure 2) shows that the reflection data are nearly perfectly interpolated even in the complex triplication zones.

The constant-velocity test results allow us to conclude that, when all the assumptions of the offset continuation theory are met, we can easily accomplish an accurate interpolation. In the next subsection, I deal with the more complicated case of Marmousi.

Marmousi synthetic

The famous Marmousi synthetic was modeled over a very complicated velocity and reflector structure (Versteeg, 1994). The dataset has been used in numerous studies of various seismic processing and imaging techniques. Figure 6 shows the near and far common-offset gathers from the Marmousi dataset. The structure of the reflection events is extremely complex and contains multiple triplications and diffractions.

To test the proposed interpolation method, I set the goal of interpolating the missing near offsets in the Marmousi dataset. Additionally, I attempted to interpolate intermediate shot gathers so that all common-midpoint gathers receive the same offset fold. In the original dataset, both receiver and shot spacing are equal to 25 meters, which creates a checkerboard pattern in the offset-midpoint plane. This acquisition pattern is typical for 2-D seismic surveys.

Interpolation of near offsets can reduce imaging artifacts in different migration methods. Ji (1995) used near-offset interpolation for accurate wavefront-synthesis migration of the Marmousi dataset. He developed an interpolation technique based on the hyperbolic Radon transform inversion. Ji's method produces fairly good results, but is significantly more expensive than the offset continuation approach explored in this paper.

Figure 7 shows the input and interpolated Marmousi data in the log-stretch frequency domain. We can see that the data in the frequency slices also have a very complicated structure. Nevertheless, the offset continuation method is able to reconstruct the missing portions of the data in a visually pleasing way. The data are not extrapolated off the sides of the common-offset gathers. This behavior is physically reasonable, because such an extrapolation would involve assumptions about unilluminated portions of the subsurface.

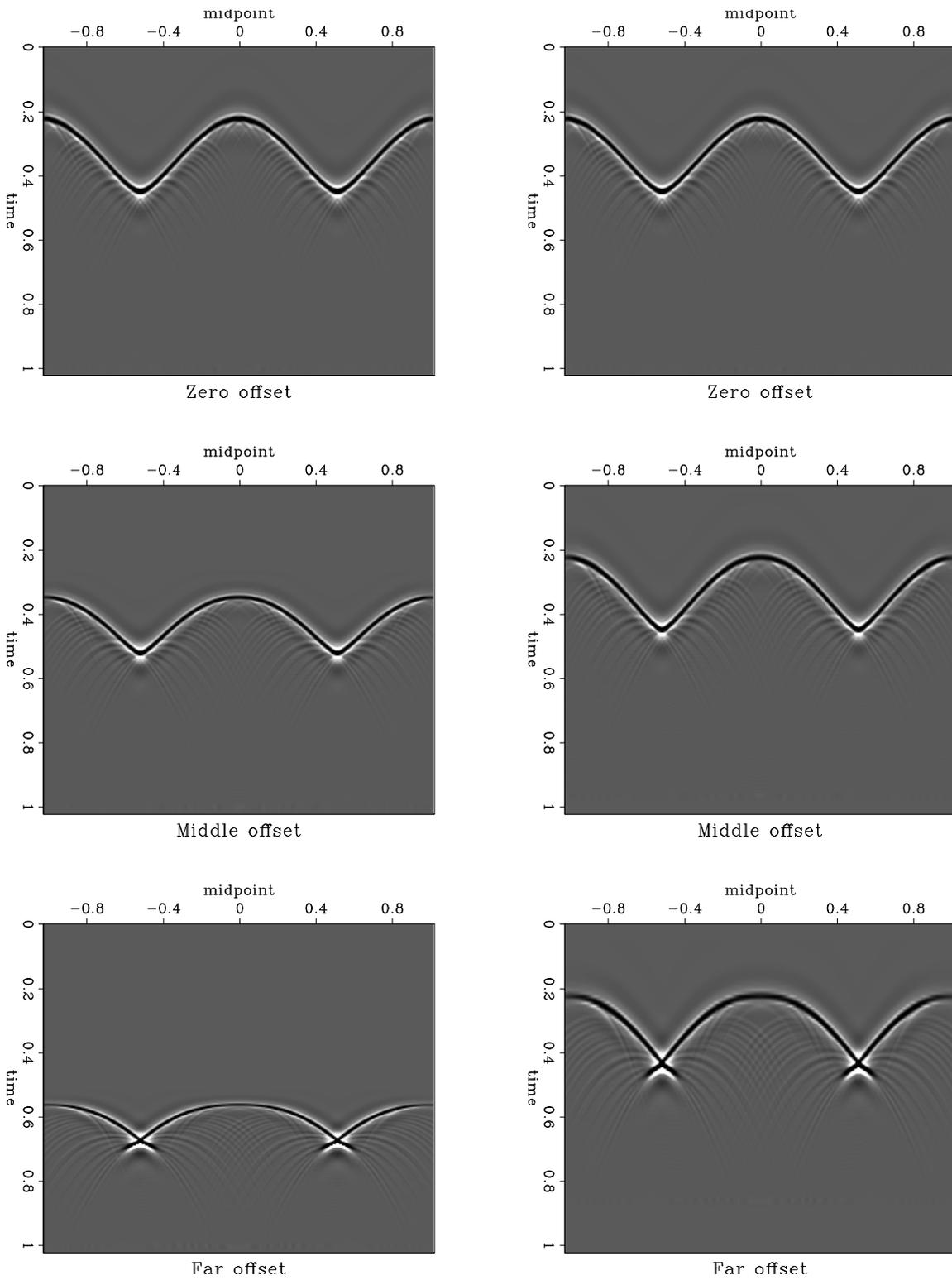


Figure 2: Prestack common-offset gathers for the constant-velocity test. Left: before NMO. Right: after NMO. Top, center, and bottom plots correspond to different offsets. `sergey2-data` [ER,M]

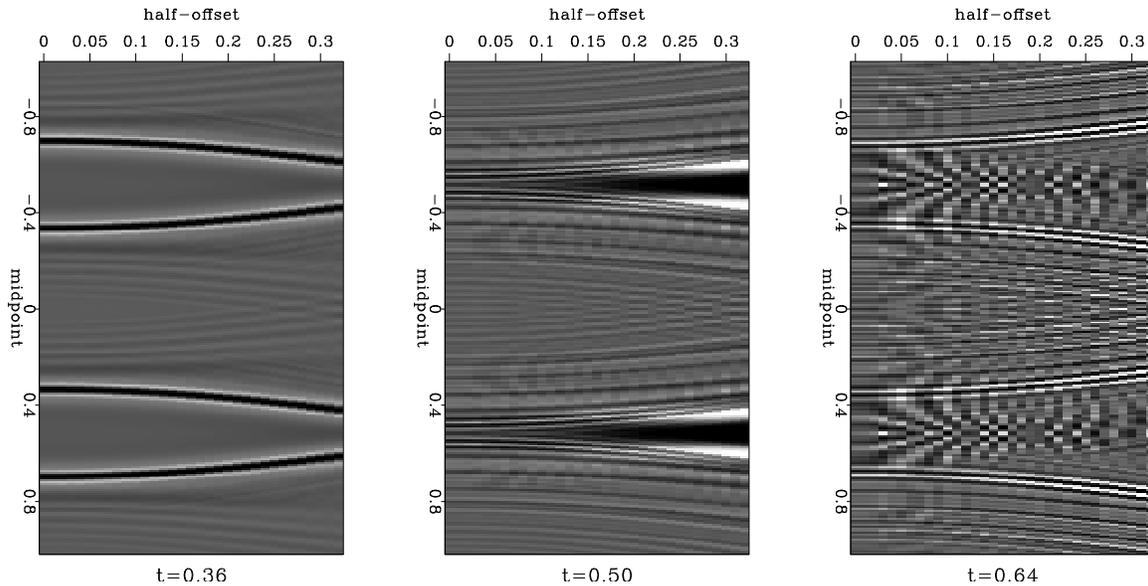


Figure 3: Time slices of the prestack data at different times (after NMO). `sergey2-tslic` [ER]

Figure 8 shows one of the shot gathers obtained after transforming the data back into time domain and resorting them into shot gathers. The positive offset part of the shot gather was reconstructed from a common receiver gather by using reciprocity. Comparing the top and bottom plots, we can see that many different events in the original shot gather are nicely extended into near offsets by the interpolation procedure.

In addition to interpolating near offsets, I have reconstructed the intermediate shot gathers in order to equalize the CMP fold. Figure 9 shows an example of an artificial shot gather created by such a reconstruction. An sample CMP gather before and after interpolation is shown in Figure 10. Examining the bottom part of the section, we can see that that the interpolation process tends to put more continuity in the near offsets than could be expected from the data. In other places, the interpolation succeeds in producing a visually pleasant result.

DISCUSSION

Early tests with synthetic models demonstrate that the offset continuation equation is a useful and efficient regularization operator for interpolating seismic reflection data. I plan to perform more tests in order to evaluate the performance of this method on real data. An extension to 3-D data is simple in theory, but it will require several modifications in the computational framework.

In the range of possible interpolation methods (Mazzucchelli et al., 1998), the offset continuation approach clearly stands on the more efficient side. The efficiency is achieved both by the small size of the finite-difference filter and by the method's ability to decompose and parallelize the method across different frequencies. Part of the efficiency gain could probably

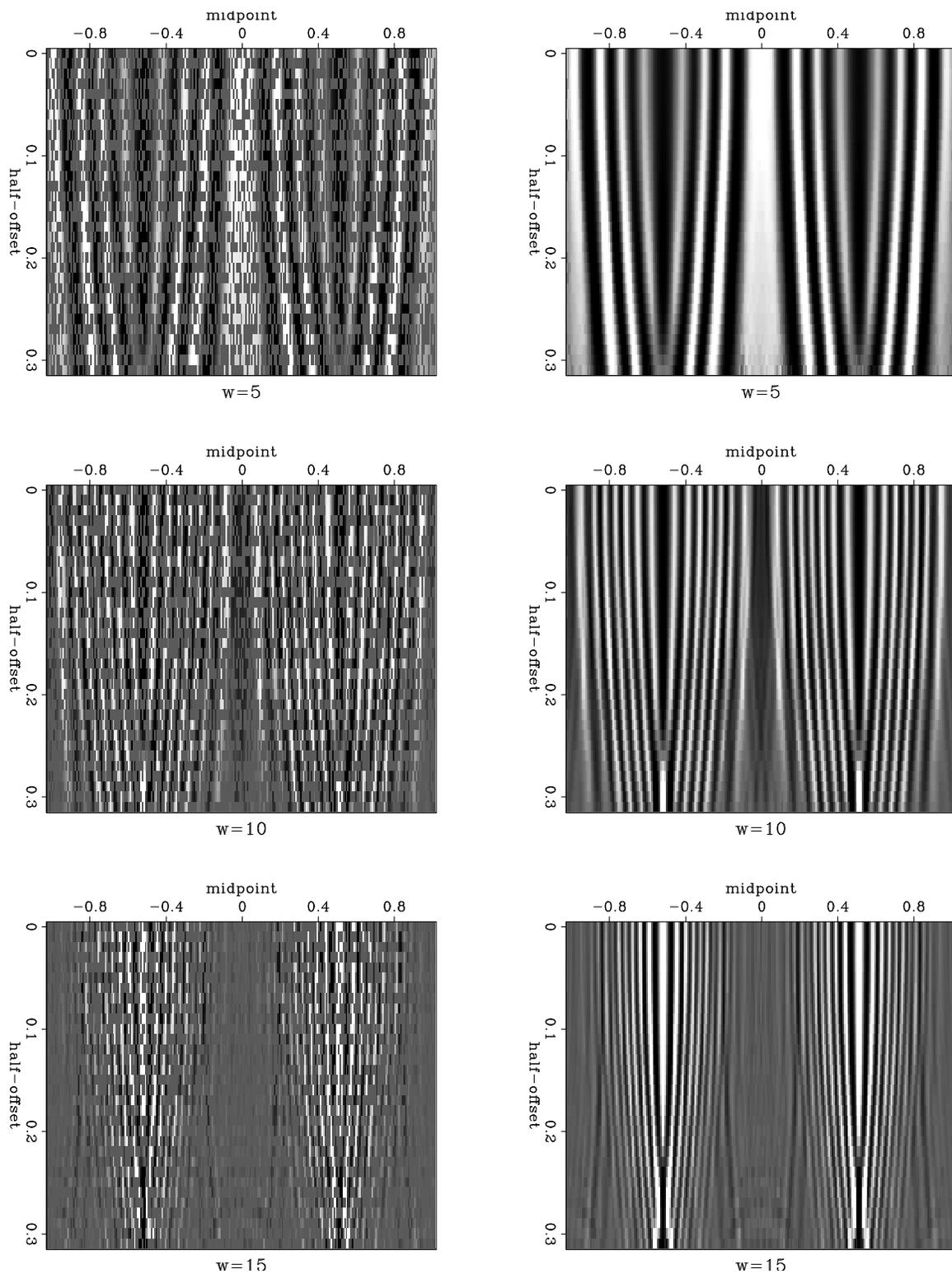


Figure 4: Interpolation in frequency slices. Left: input data (50% of the traces are randomly removed). Right: interpolation output. Top, bottom, and middle plots correspond to different frequencies. Real parts of the complex-valued data are shown. `sergey2-fslice` [ER,M]

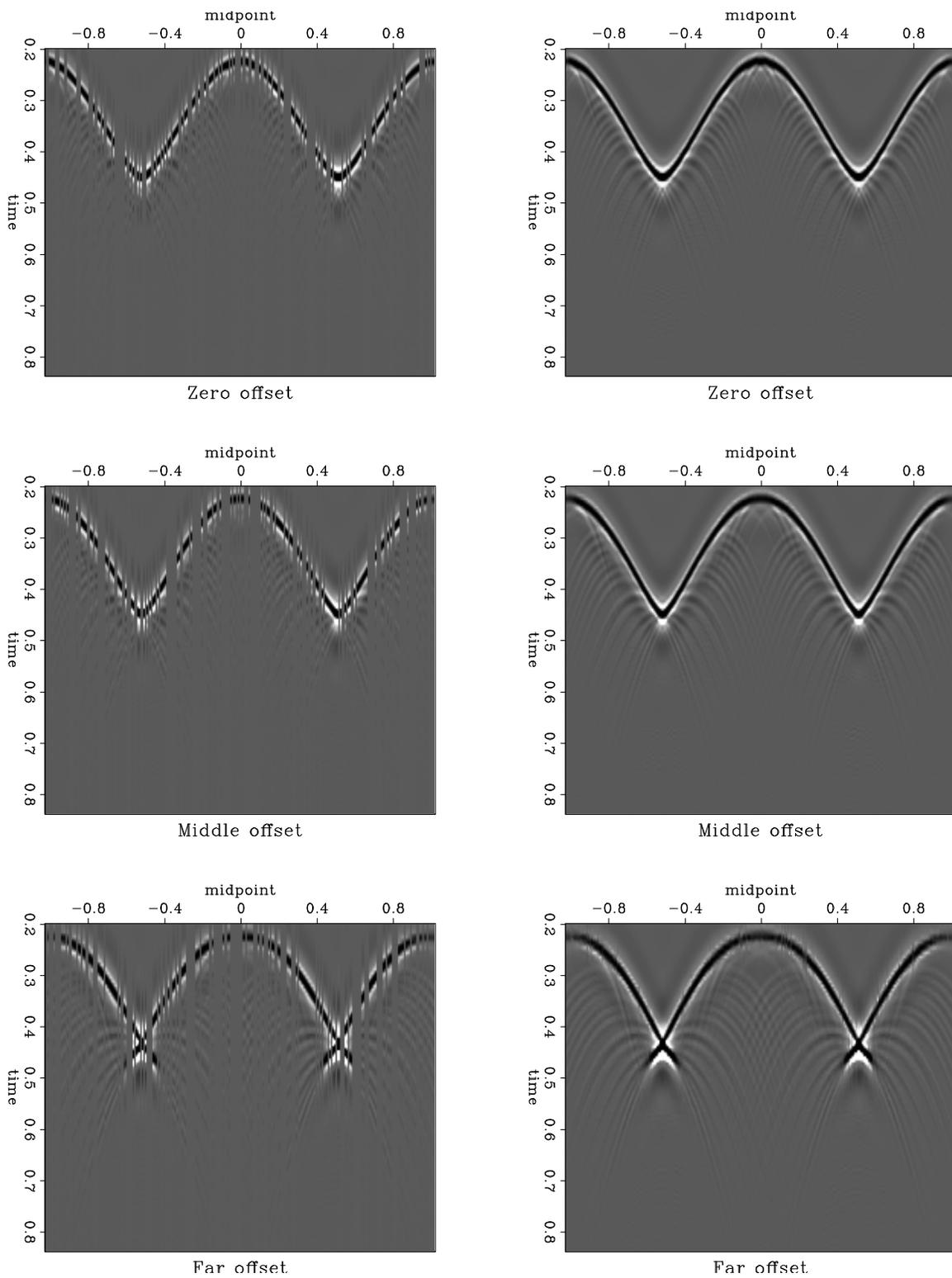


Figure 5: Interpolation in common-offset gathers. Left: input data (50% of the traces are randomly removed). Right: interpolation output. Top, center, and bottom plots correspond to different common-offset gathers. `sergey2-all` [ER,M]

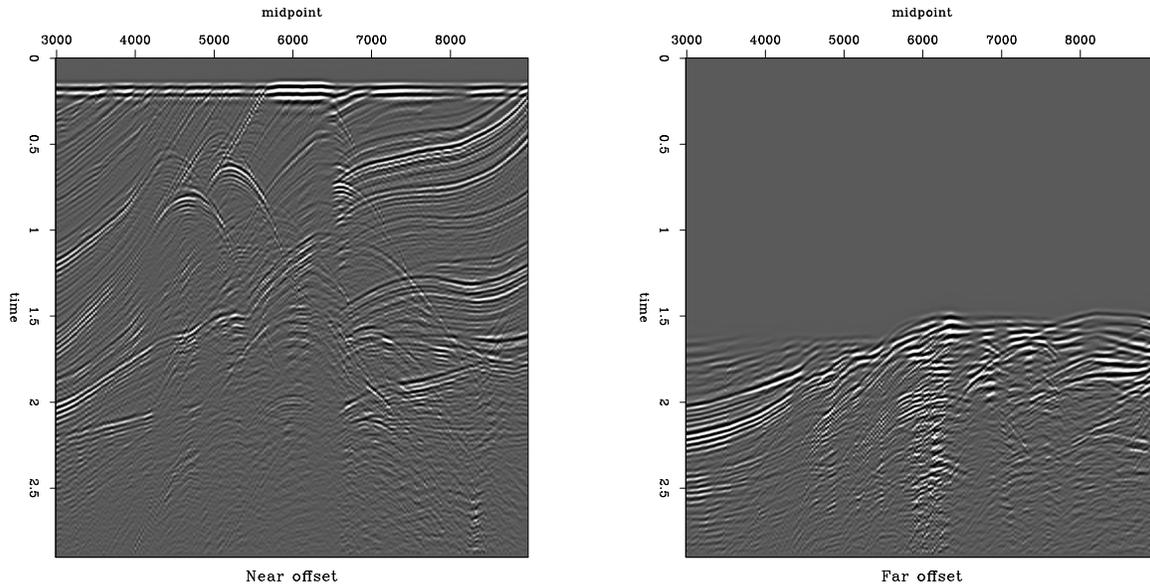


Figure 6: Common-offset gathers of the Marmousi dataset. Left: near offset. Right: far offset. `sergey2-marm` [ER]

be sacrificed for achieving more accurate results. Here are some interesting ideas one could try:

- Instead of fixing the offset continuation filter in a data-independent way, one could estimate some of its coefficients from the data. In particular, the second term in equation (3) can be varied to better account for the effects of variable velocity and amplitude variation with offset. Theoretical extensions of offset continuation to the variable velocity case were studied by Hong et al. (1997) and Luo et al. (1999).
- Formulating the problem in the pre-NMO domain would allow us to consider several velocities by convolving several continuation filters. This could be an appropriate approach for interpolating both primary and multiple reflections.
- Missing data interpolation problems can be greatly accelerated by preconditioning (Fomel, 1997; Fomel et al., 1997). Finding an appropriate preconditioning for the offset continuation method is an open problem. The non-stationary nature of the continuation filter make this problem particularly challenging.

I plan to devote my remaining time at the Stanford Exploration Project to investigating these fascinating ideas.

REFERENCES

Bagaini, C., and Spagnolini, U., 1993, Common shot velocity analysis by shot continuation operator: 63rd Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 673–676.

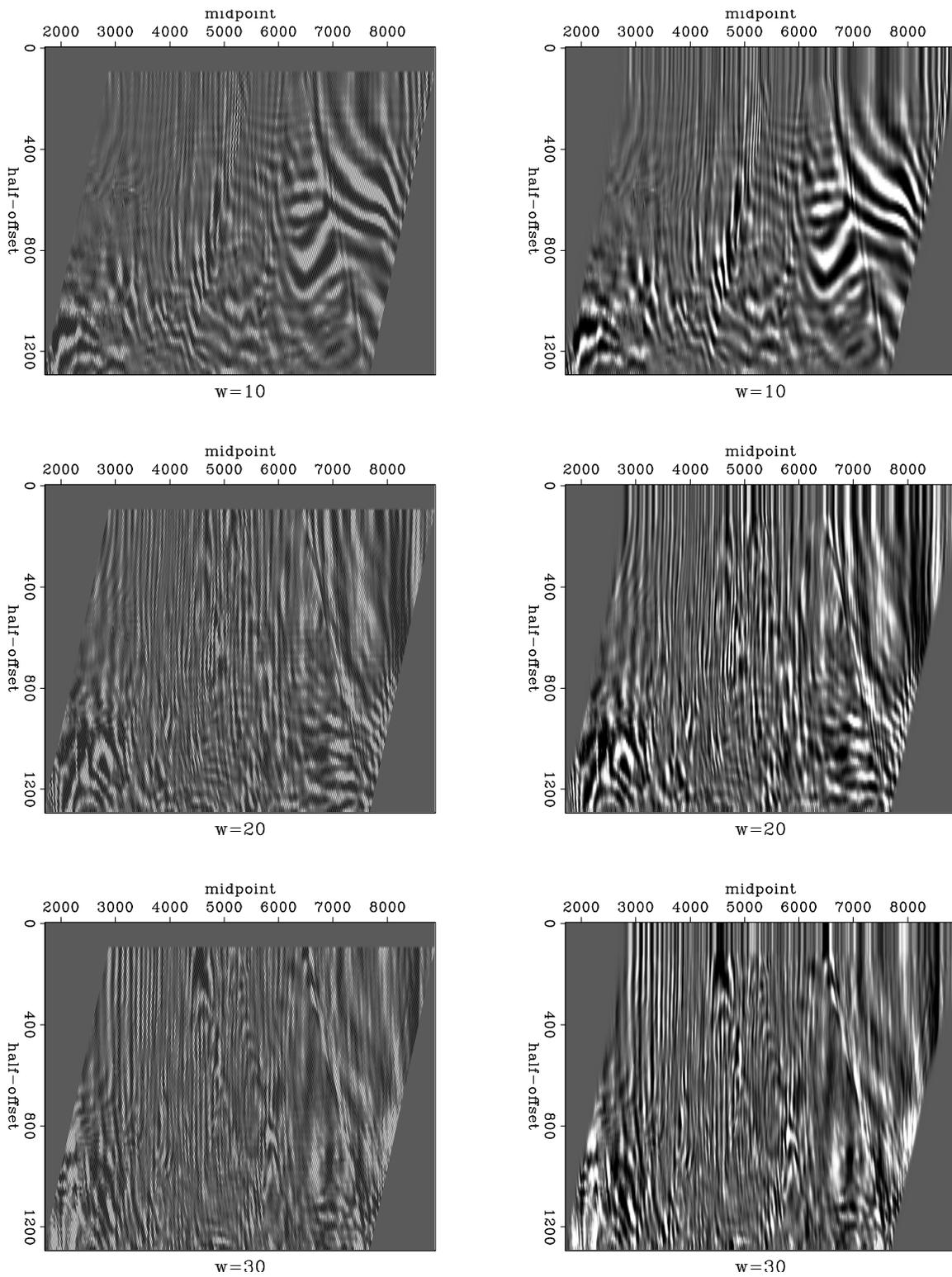


Figure 7: Interpolation of the Marmousi dataset in frequency slices. Left: input data. Right: interpolation output. Top, center, and bottom plots correspond to different frequencies. Real parts of the complex-valued data are shown. `sergey2-mslice` [CR,M]

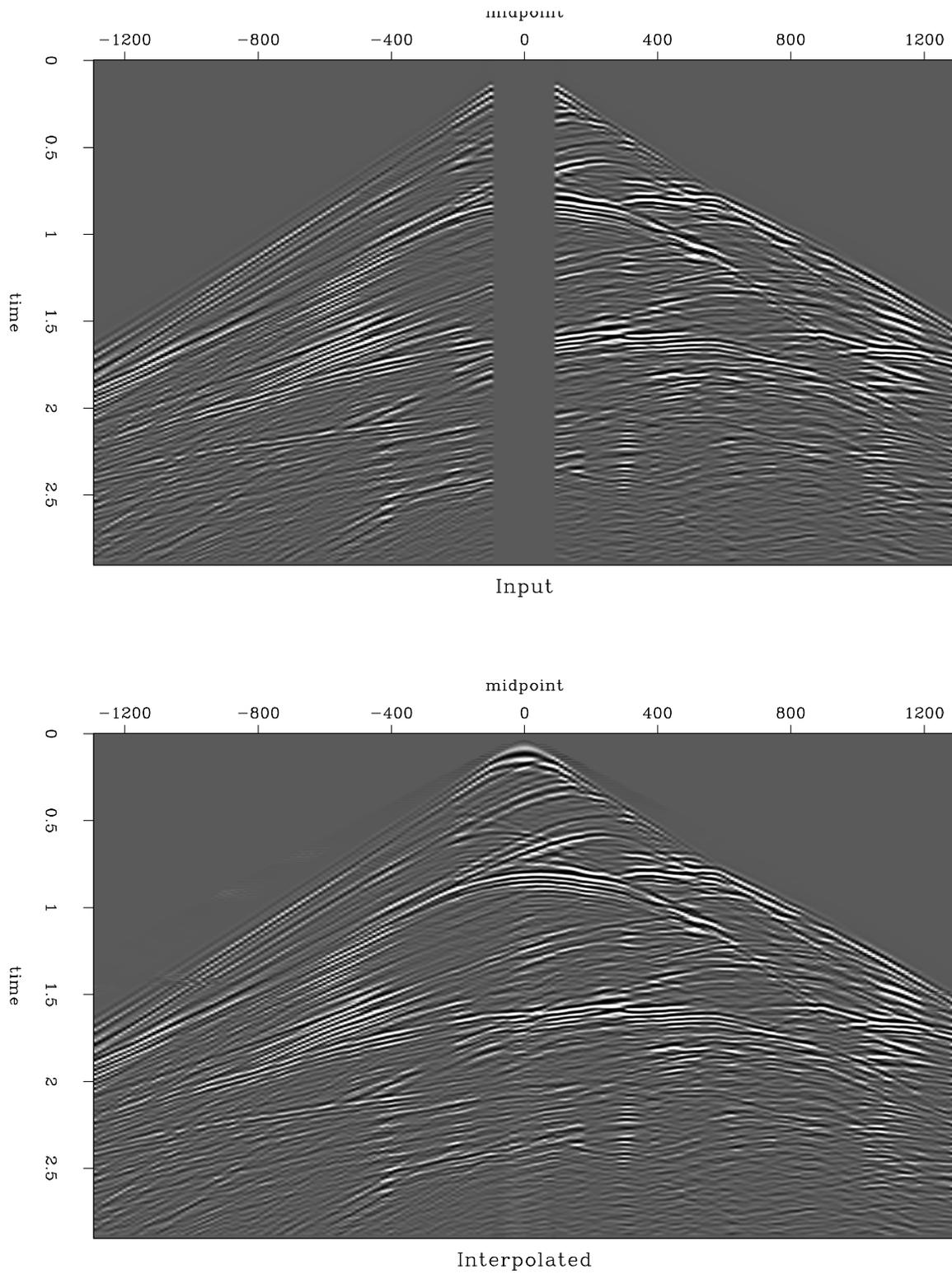


Figure 8: Interpolation of near offsets in a Marmoussi shot gather. The shot position is 4500 m. Top: input data. Bottom: interpolation output. `sergey2-mshot` [CR,M]

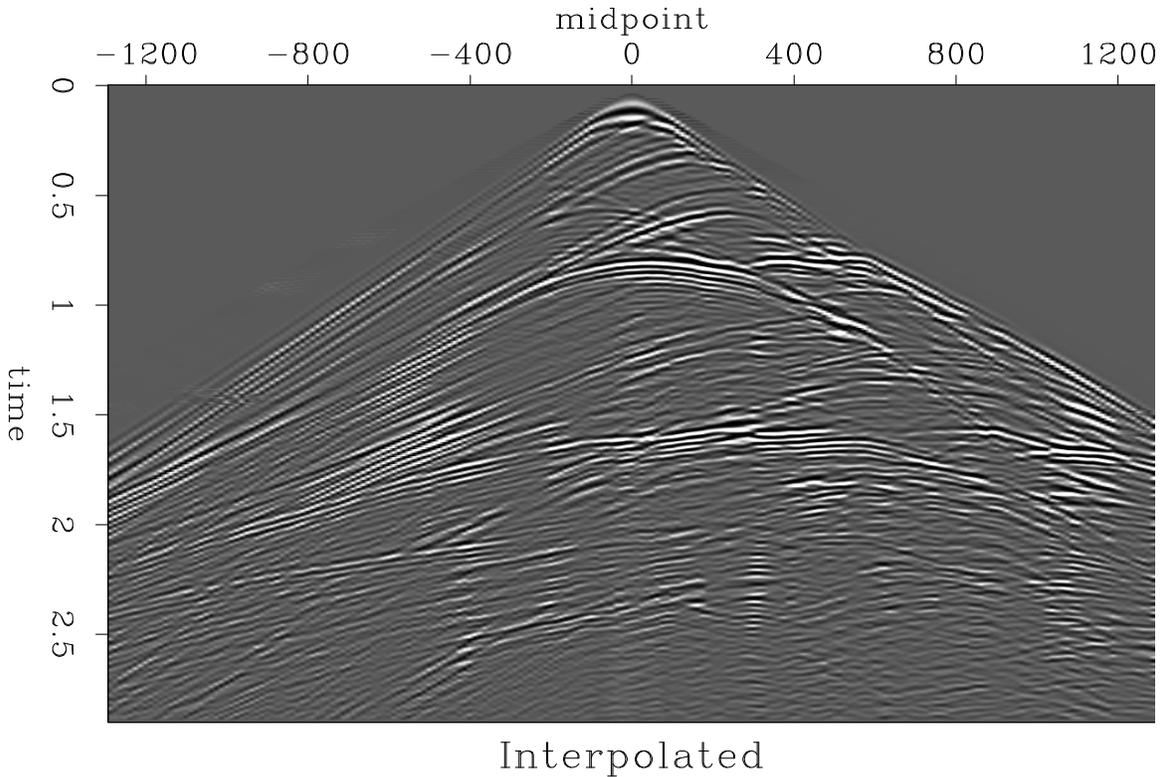


Figure 9: This shot gather at 4525 m is a result of data interpolation. sergey2-mishot [CR]

Bagaini, C., and Spagnolini, U., 1996, 2-D continuation operators and their applications: *Geophysics*, **61**, no. 06, 1846–1858.

Bagaini, C., Spagnolini, U., and Paziienza, V. P., 1994, Velocity analysis and missing offset restoration by prestack continuation operators: 64th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1549–1552.

Biondi, B., and Palacharla, G., 1996, 3-D prestack migration of common-azimuth data: *Geophysics*, **61**, no. 06, 1822–1832.

Biondi, B., Fomel, S., and Chemingui, N., 1998, Azimuth moveout for 3-D prestack imaging: *Geophysics*, **63**, no. 2, 574–588.

Biondi, B. L., 1999, 3-D Seismic Imaging: Stanford Exploration Project.

Bolondi, G., Loinger, E., and Rocca, F., 1982, Offset continuation of seismic sections: *Geophys. Prosp.*, **30**, no. 6, 813–828.

Briggs, I. C., 1974, Machine contouring using minimum curvature: *Geophysics*, **39**, no. 1, 39–48.

Chemingui, N., and Biondi, B., 1994, Coherent partial stacking by offset continuation of 2-D prestack data: SEP-**82**, 117–126.

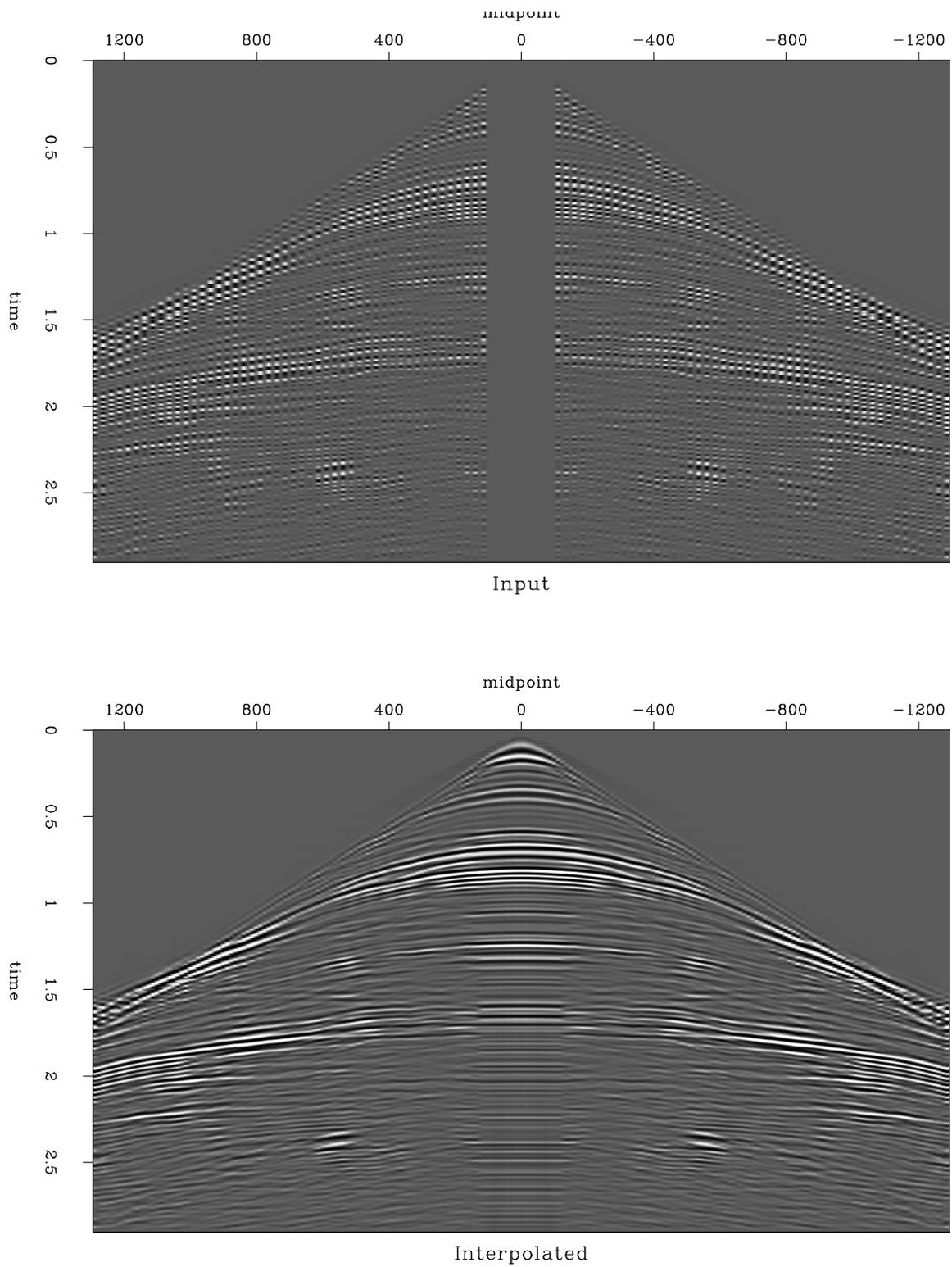


Figure 10: Interpolation of near and intermediate offsets in a Marmousi CMP gather. The midpoint position is 4500 m. Top: input data. Bottom: interpolation output. `sergey2-mcmp` [CR,M]

- Chemingui, N., 1999, Imaging irregularly sampled 3D prestacked data: Ph.D. thesis, Stanford University.
- Claerbout, J. F., 1980, Restoration of missing data by least squares optimization: SEP-25, 1-16.
- Claerbout, J. F., 1981, Missing data: Wanted, good convergence in a few steps: SEP-26, 23-38.
- Claerbout, J. F., 1992, Earth Soundings Analysis: Processing Versus Inversion: Blackwell Scientific Publications.
- Claerbout, J., 1999, Geophysical estimation by example: Environmental soundings image enhancement: Stanford Exploration Project, <http://sepwww.stanford.edu/sep/prof/>.
- Clapp, R. G., Fomel, S., and Claerbout, J., 1997, Solution steering with space-variant filters: SEP-95, 27-42.
- Clapp, R. G., Fomel, S., Crawley, S., and Claerbout, J. F., 1999, Directional smoothing of non-stationary filters: SEP-100, 197-209.
- Crawley, S., 1999, Interpolation with smoothly nonstationary prediction-error filters: SEP-100, 181-196.
- Deregowski, S. M., and Rocca, F., 1981, Geometrical optics and wave theory of constant offset sections in layered media: Geophys. Prosp., 29, no. 3, 374-406.
- Deregowski, S. M., 1986, What is DMO: First Break, 4, no. 7, 7-24.
- Fomel, S., and Biondi, B., 1995, The time and space formulation of azimuth moveout: SEP-84, 25-38.
- Fomel, S., and Bleistein, N., 1996, Amplitude preservation for offset continuation: Confirmation for Kirchhoff data: SEP-92, 219-227.
- Fomel, S., Bleistein, N., Jaramillo, H., and Cohen, J. K., 1996, True amplitude DMO, offset continuation and AVA/AVO for curved reflectors: 66th Ann. Internat. Meeting, Soc. Expl. Geophys., Expanded Abstracts, 1731-1734.
- Fomel, S., Clapp, R., and Claerbout, J., 1997, Missing data interpolation by recursive filter preconditioning: SEP-95, 15-25.
- Fomel, S. B., 1994, Kinematically equivalent differential operator for offset continuation of seismic sections: Russian Geology and Geophysics, 35, no. 9, 122-134.
- Fomel, S., 1995a, Amplitude preserving offset continuation in theory Part 1: The offset continuation equation: SEP-84, 179-198.
- Fomel, S., 1995b, Amplitude preserving offset continuation in theory part 2: Solving the equation: SEP-89, 109-132.

- Fomel, S., 1997, On model-space and data-space regularization: A tutorial: SEP-94, 141–164.
- Fomel, S., 1999, Plane wave prediction in 3-D: SEP-102, 101–110.
- Fomel, S., 2000, Helical preconditioning and splines in tension: SEP-103, 289–301.
- Gardner, G. H. F., and Canning, A., 1994, Effects of irregular sampling on 3-D prestack migration: 64th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1553–1556.
- Hale, D., Ed. **DMO processing**. Society Of Exploration Geophysicists, 1995.
- Hong, L., Xingyao, Y., Youming, L., Changchun, Y., and McMechan, G. A., 1997, The offset continuation in complex medium: 67th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1901–1904.
- Ji, J., 1995, Sequential seismic inversion using plane-wave synthesis: Ph.D. thesis, Stanford University.
- Luo, M., Li, Y., and Liu, H., 1999, Amplitude preserving offset continuation in lateral varied media: 69th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1933–1936.
- Mazzucchelli, P., and Rocca, F., 1999, Regularizing land acquisitions using shot continuation operators: effects on amplitudes: 69th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1995–1998.
- Mazzucchelli, P., Rocca, F., Spagnolini, U., and Spitz, S., 1998, Wavefield interpolation - Continuation or prediction filter techniques?: 61st Mtg. Eur. Assoc. Expl Geophys., Extended Abstracts, Session:02–51.
- Salvador, L., and Savelli, S., 1982, Offset continuation for seismic stacking: Geophys. Prosp., **30**, no. 6, 829–849.
- Schwab, M., 1993, Shot gather continuation: SEP-77, 117–130.
- Spagnolini, U., and Opreni, S., 1996, 3-D shot continuation operator: 66th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 439–442.
- Spitz, S., 1991, Seismic trace interpolation in the f-x domain: Geophysics, **56**, no. 6, 785–794.
- Stovas, A. M., and Fomel, S. B., 1996, Kinematically equivalent integral DMO operators: Russian Geology and Geophysics, **37**, no. 2, 102–113.
- Symes, W. W., and Carazzone, J. J., 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, no. 5, 654–663.
- Symes, W. W., 1999, All stationary points of differential semblance are asymptotic global minimizers: Layered acoustics: SEP-100, 71–92.
- Tarantola, A., 1987, Inverse problem theory: Elsevier.

- Thorson, J. R., 1981, Restoration of missing data by parsimony in the frequency domain: SEP-26, 1-22.
- van Dedem, E. J., and Verschuur, D. J., 1998, 3-D surface-related multiple elimination and interpolation: 68th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1321-1324.
- Versteeg, R., 1994, The marmousi experience: Velocity model determination on a synthetic complex data set: The Leading Edge, **13**, no. 09, 927-936.

Huber solver versus IRLS algorithm for quasi L1 inversion

*Antoine Guitton*¹

ABSTRACT

The l^1 problem can be approximated using either Iteratively Reweighted Least Squares methods (IRLS) with an appropriate weighting function or the Huber misfit function with an appropriate solver. Comparisons of both methods show that they give very similar results as they handle problems suited to l^1 properly. In addition, the nonlinear scheme, using the Huber norm, generally converges better than the linear IRLS algorithms on velocity-stack inversion. Finally, the Huber solver is far easier to operate because it is sensitive to fewer parameters than IRLS.

INTRODUCTION

Robust error measures such as the l^1 norm have a number of uses in geophysics. As measures of data misfit, they show considerably less sensitivity to large measurement errors than does the least-squares (l^2) measure. In a recent work, Guitton and Symes (1999) show that a possible approximation of the l^1 norm is viable using the Huber misfit function (Huber, 1973). In particular, their work on velocity inversion, using the hyperbolic Radon transform, suggests that (1) the Huber function gives far more robust model estimates than does least squares, (2) its minimization using a standard quasi-Newton method is comparable in computational cost to least-squares estimation using conjugate gradient iteration, and (3) the result of Huber data fitting is stable over a wide range of choices for the $l^2 \rightarrow l^1$ threshold and total number of quasi-Newton steps. For the same geophysical problem, Nichols (1994) and Hugonnet (1998) use a different implementation of the hybrid l^2 - l^1 norm, using an IRLS algorithm with a thoroughly chosen weighting function. Both strategies (IRLS and the Huber solver) show similar results, but at this stage, no direct comparisons between the two methods have been carried out.

This paper will compare the efficiency of these two robust solvers with synthetic and real data. First, I will briefly review both methods and describe algorithmic issues. Then, I will compare the Huber solver and IRLS performing velocity inversion as suggested by Thorson and Claerbout (1985), Nichols (1994), and Lumley et al. (1994).

¹email: antoine@sep.stanford.edu

IRLS AND THE HUBER SOLVER, AN OVERVIEW

Minimizing the Huber function

In the work reported here, I use a hybrid l^1 - l^2 error measure proposed by Huber (Huber, 1973):

$$M_\epsilon(r) = \begin{cases} \frac{r^2}{2\epsilon}, & 0 \leq |r| \leq \epsilon \\ |r| - \frac{\epsilon}{2}, & \epsilon < |r| \end{cases}$$

I will call $\sum_{i=1}^N M_\epsilon(r_i)$ the *Huber misfit function* H_ϵ , or Huber function for short (Figure 1). Note that the Huber function is smooth near zero residual, and weights small residuals by the mean square. It is reasonable to suppose that the Huber function, while maintaining robustness against large residuals, is easier to minimize than l^1 . The parameter ϵ , which controls the limit between l^1 and l^2 , is called the Huber threshold.

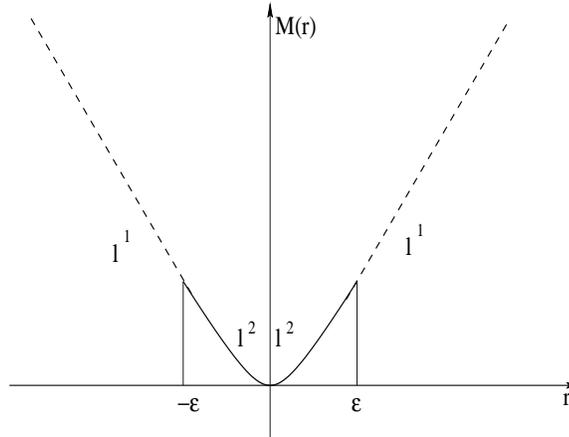


Figure 1: Error measure proposed by Huber (Huber, 1973). The upper part above ϵ is the l^1 norm, while the lower part is the l^2 norm.

`antoine2-huber` [NR]

The problem is this: given some observed data \mathbf{d}_{obs} , we want to find the best model \mathbf{m} that explains the data *via* the operator \mathbf{H} . This may be posed in terms of an inverse problem leading to the minimization of the function

$$f(\mathbf{m}) = E(\mathbf{H}\mathbf{m} - \mathbf{d}_{\text{obs}}), \quad (1)$$

where E is an error measure function. As discussed above, I will use the Huber function and try to minimize

$$f(\mathbf{m}) = H_\epsilon(\mathbf{H}\mathbf{m} - \mathbf{d}_{\text{obs}}). \quad (2)$$

We seek to find the stationary point \mathbf{m}^* of the function f . This solution point satisfies $\mathbf{f}'(\mathbf{m}^*) = \mathbf{0}$. This is a nonlinear system of equations, and from Taylor expansion we have

$$\mathbf{f}'(\mathbf{m} + \delta\mathbf{m}) \simeq \mathbf{f}'(\mathbf{m}) + \mathbf{f}''(\mathbf{m})\delta\mathbf{m}$$

if $\|\delta\mathbf{m}\|$ is sufficiently small. The *Newton's method* consists in finding $\delta\mathbf{m}$ such that

$$\mathbf{H}\delta\mathbf{m} = -\mathbf{f}'(\mathbf{m}) \text{ with } \mathbf{H} = \mathbf{f}''(\mathbf{m}),$$

and computing the next iterate by

$$\mathbf{m}_{n+1} = \mathbf{m}_n + \alpha_n \delta \mathbf{m}_n \quad (3)$$

where α_n is a steplength given by a Line Search algorithm. The general process of the program is then

1. compute the gradient $\mathbf{f}'(\mathbf{m})$
2. compute $\delta \mathbf{m}_n = -\mathbf{H}^{-1} \mathbf{f}'(\mathbf{m})$
3. compute α_n using a line search
4. update the solution $\mathbf{m}_{n+1} = \mathbf{m}_n + \alpha_n \delta \mathbf{m}_n$
5. update the Hessian \mathbf{H}
6. go to 1.

Because the Huber function is not twice continuously differentiable, the Hessian is not computed directly but approximated using a limited Memory BFGS update (Guitton, 2000), as proposed by Nocedal (1980) and Liu and Nocedal (1989). I have implemented for the line search a More and Thuente (More and Thuente, 1994) algorithm, which ensures sufficient decrease for the function f and obeys curvature conditions (the so-called *Wolfe conditions*, Kelley (1999)), thus guaranteeing that a quasi-Newton update is possible. The steplength $\alpha_n = 1$ is always tried first, saving significant computation. These choices lead to good performances for both convergence rate and computation cost. I call “Huber solver” the algorithm detailed above when used with the Huber norm.

The Iteratively Reweighted Least Square method

The IRLS implementation of the hybrid l^2 - l^1 norm differs greatly from the Huber solver. In this section, I follow quite closely what Nichols (1994) and Darche (1989) suggested in previous reports. The main idea is that instead of minimizing the simple l^2 norm, we can choose to minimize a weighted residual. The objective function becomes

$$f(\mathbf{m}) = \|\mathbf{W}\mathbf{H}\mathbf{m} - \mathbf{W}\mathbf{d}_{obs}\|_2. \quad (4)$$

A difference between the Huber solver and IRLS clearly arises here since we still solve a least-square problem with IRLS and not with the Huber solver. The IRLS method weights residuals within a linear l^2 framework and Huber uses either l^2 or l^1 following the residual with a nonlinear update. A particular choice for \mathbf{W} will lead to the minimization of the l^2 - l^1 norm, however. In this paper, I choose

$$w_{ii} = \frac{1}{[1 + (r_{ii}/\epsilon)^2]^{1/4}}, \quad (5)$$

with $r_i = \mathbf{H}\mathbf{m}_i - \mathbf{d}_i$ and ϵ a damping parameter. With this particular choice of \mathbf{W} , minimizing f is equivalent to minimizing (Bube and Langan, 1997)

$$J(\mathbf{m}) = \sum_{i=1}^n j(r_i) = \sum_{i=1}^n \sqrt{1 + (r_i/\epsilon)^2} - 1. \quad (6)$$

For any given residual r_i , we notice that

$$j(r) \approx \begin{cases} \frac{1}{2}(r/\epsilon)^2 & \text{for } |r| \text{ small} \\ |r|/\epsilon & \text{for } |r| \text{ large.} \end{cases}$$

Hence, we obtain a l^1 treatment of large residuals and a l^2 treatment of small residuals. Darche (1989) gives a practical value for ϵ that can be also used for the Huber threshold:

$$\epsilon = \frac{\max|\mathbf{d}|}{100}.$$

Two different IRLS algorithms are suggested in the literature (Scales and Gersztenkorn, 1988). The first method consists of solving successive l^2 problems with a constant weight and taking the last solution obtained to recompute the weight for the next l^2 problem. The second method consists of recomputing the weights at each iteration (in fact, we need only a few iterations of the conjugate gradient scheme before calculating the new weight), solving small piecewise linear problems. The IRLS algorithms converge if each minimization reaches a minimum for a constant weight (Bube and Langan, 1997). This is the case for the first method described above, but it is also true for the second method since each starting point of the next iteration is the ending point of the previous solving piecewise linear problems. For practical reasons, we utilize the second method. Indeed, for most geophysical problems, the complete solution of the l^2 problem with a constant weight would lead to prohibitive costs. The general process in the program is as follows:

1. compute the current residual $\mathbf{r} = \mathbf{H}\mathbf{m} - \mathbf{d}$
2. compute the weighting operator \mathbf{W} using \mathbf{r}
3. solve the weighted least-squares problem (equation 4) using a Conjugate Gradient algorithm
4. go to first step

We do not detail the Conjugate Gradient step here. For more details, the reader may refer to Paige and Saunders (1982). In our implementation, we control the restarting schedule of the iterative scheme with one parameter that governs the periodic computing of the weight. The Conjugate Gradient is reset for each new weighting function, meaning that the first iteration of each new least-squares problem (for each new weight) is a steepest descent step. In addition, the last solution \mathbf{m} of the previous least-squares problem is used to compute the new residual and the new weighting matrix (steps 1 and 2 in the IRLS algorithm). In the following examples, except when indicated, we recompute the weight every five iterations. In the following sections, I call “IRLS” the IRLS algorithm described above with the weighting function in equation (5).

COMPARING IRLS AND HUBER ON A GEOPHYSICAL PROBLEM

We now compare IRLS and the Huber solver doing velocity-stack inversion. This method was first introduced by Thorson and Claerbout (1985) and completed with an IRLS algorithm by Nichols (1994). Guitton and Symes (1999) introduced the Huber misfit function with a specific non-linear solver as an alternative to IRLS. The velocity estimation is made using inverse methods to minimize the function

$$f(\mathbf{m}) = E(\mathbf{H}\mathbf{m} - \mathbf{d}),$$

where E is an error measure function, \mathbf{H} is the modeling operator that transforms the model space \mathbf{m} (velocity domain) into the data space \mathbf{d} (CMP gathers). The adjoint operator (\mathbf{H}^\dagger) is the well-known hyperbolic Radon transform operator. The forward operation is

$$d(t, x) = \sum_{s=s_{min}}^{s_{max}} w_o m(\tau = \sqrt{t^2 - s^2 x^2}, s), \quad (7)$$

and the adjoint transformation becomes

$$m(\tau, s) = \sum_{x=x_{min}}^{x_{max}} w_o d(t = \sqrt{\tau^2 + s^2 x^2}, x), \quad (8)$$

where w_o is a weighting function (Claerbout and Black, 1997). We now solve the inverse problem using both the Huber solver and IRLS on synthetic and real data. Note that in the following sections, I assume that when $\epsilon = \max|\mathbf{d}|/100$, we nearly solve the l^1 problem for both the Huber solver and IRLS.

Tests on synthetic data: spiky events

Figure 2 shows the input data for the first synthetic example plus the l^2 inversion result; it demonstrates the sensitivity of least-squares to outliers: some ellipses appear in the model space and the data space is infested with noise. Figures 3 and 4 show a comparison between IRLS and the Huber solver (same clip applied on the data). The damping parameter for IRLS and the Huber threshold were chosen as suggested above ($\epsilon = \max|\mathbf{d}|/100$). Both results are very comparable since we get rid of the spikes in the data space (Figure 3). In addition, the five expected velocity events, associated with the five input hyperbolas, are clearly identifiable in the model space (Figure 4). Both results are very conclusive since the spikes have been correctly “ l^1 handled”.

Tests on synthetic data: groundroll

In this example, we contaminated the data with a low frequency/low velocity linear event that simulates, to some extent, groundroll (Figure 5, left). Because the operator \mathbf{A} models

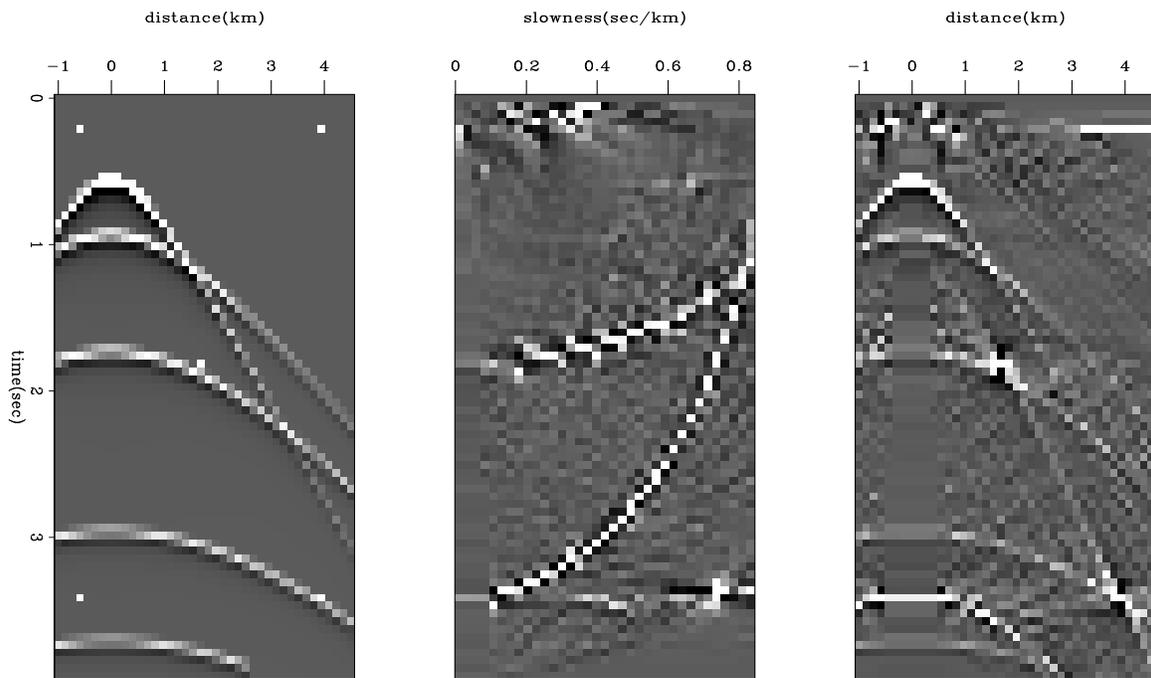


Figure 2: From left to right: 1) Input data. 2) velocity domain, l^2 inversion result. 3) remodeled data from the l^2 result `antoine2-gr-spiky-L2-HUBER` [ER]

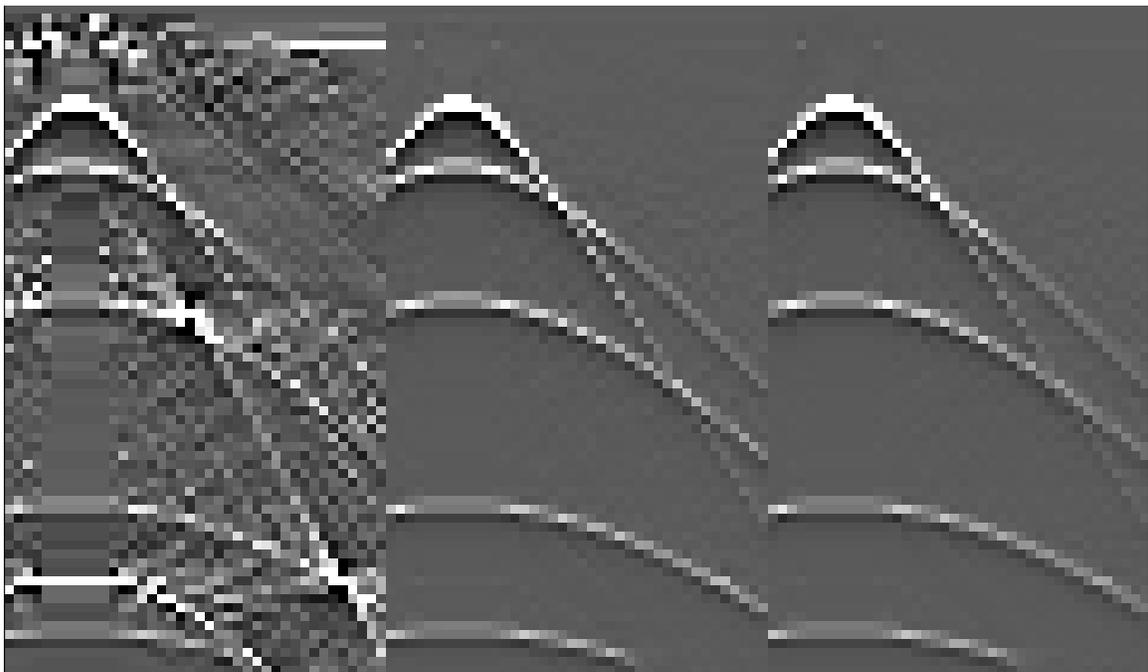


Figure 3: From left to right: 1) Remodeled data with l^2 . 2) Remodeled data with the Huber solver. 3) Remodeled data using IRLS `antoine2-comp-data-spiky-0.04` [ER]

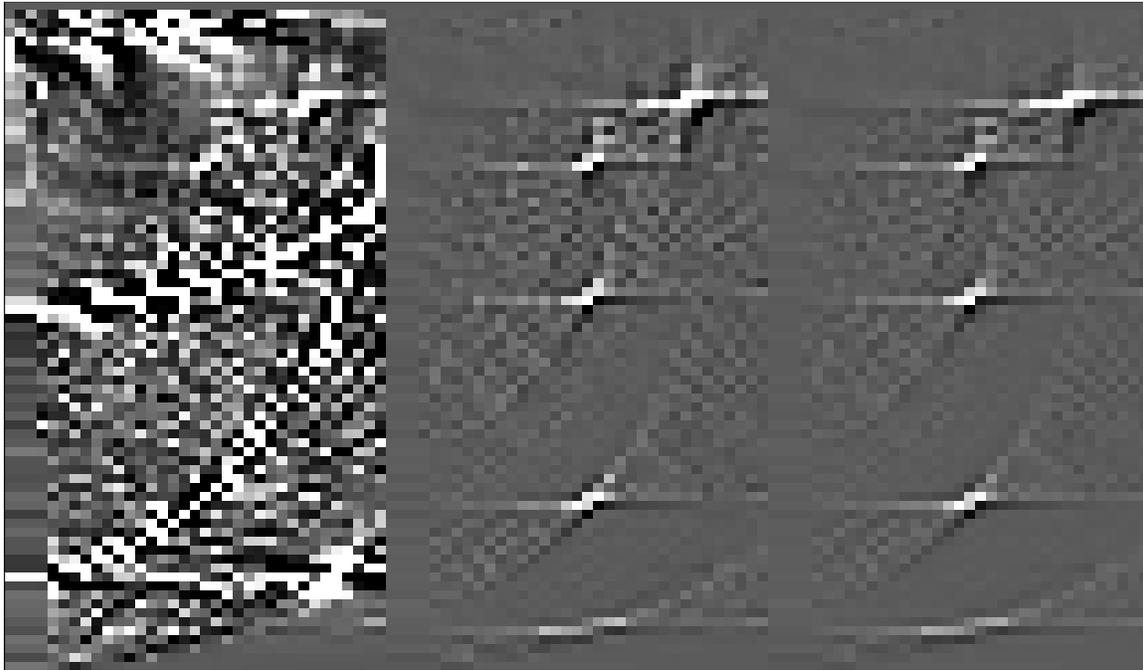


Figure 4: From left to right: 1) l^2 velocity space. 2) Huber velocity space. 3) IRLS velocity space `antoine2-comp-mod-spiky-0.04` [ER]

only hyperbolas, the simulated groundroll introduces some inconsistent data. It is therefore interesting to test the sensitivity of our inversion schemes (l^2 and l^1) for this new problem. The l^2 inversion of the data leads to a fairly noisy model where the five expected events are difficult to pick (Figure 5). However, without groundroll and only a few artifacts, the remodeled l^2 data does not display so dramatic effects. With a “ l^1 ” inversion, the model space is more accurately resolved, as shown in Figure 7. Again, there are few differences between IRLS and the Huber solver, and it looks like the “ l^1 ” norm copes more easily with inconsistent events than l^2 (groundroll in this case).

Tests on real data

Figure 8 shows the input for the real data example. These data are contaminated with a low velocity/low frequency noise crossing the main signal. The Huber threshold and the damping parameter for IRLS are equal and obey the empirical choice that we gave above. Figure 8 details the least-squares result. The velocity panel (middle) shows some horizontal stripes that severely blur the main velocity fairway; in addition, because no strong event appears below 1s., a poor resolution is achieved. In regard to the l^2 result, we might note that the remodeled data (right panel of Figure 8) exhibit some artifacts in the top part that shows the limitations of the least-squares inversion in this case. Figure 10 displays the “ l^1 ” inversion results with the Huber solver and IRLS. Again, because we are able to pick up a velocity fairway all the way down, both results are very comparable. Moreover, the remodeled data (Figure 9) show fewer

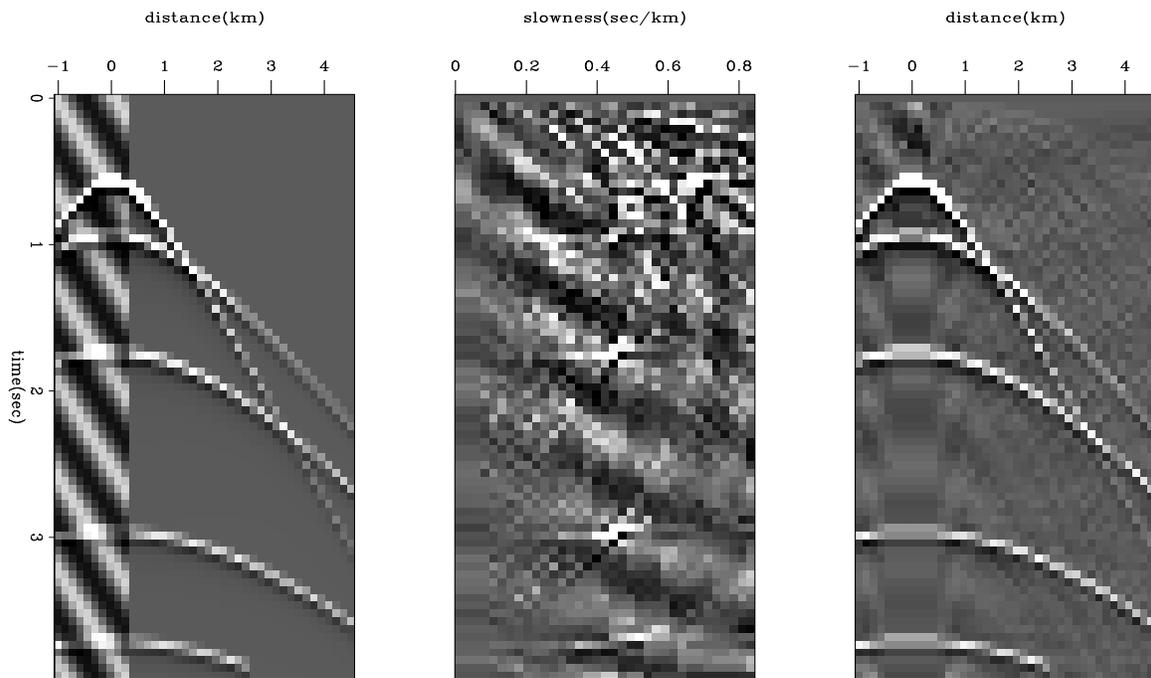


Figure 5: From left to right: 1) Input data. 2) velocity domain, l^2 inversion result. 3) remodeled data from the l^2 result `antoine2-gr-freq30-L2-HUBER` [ER]

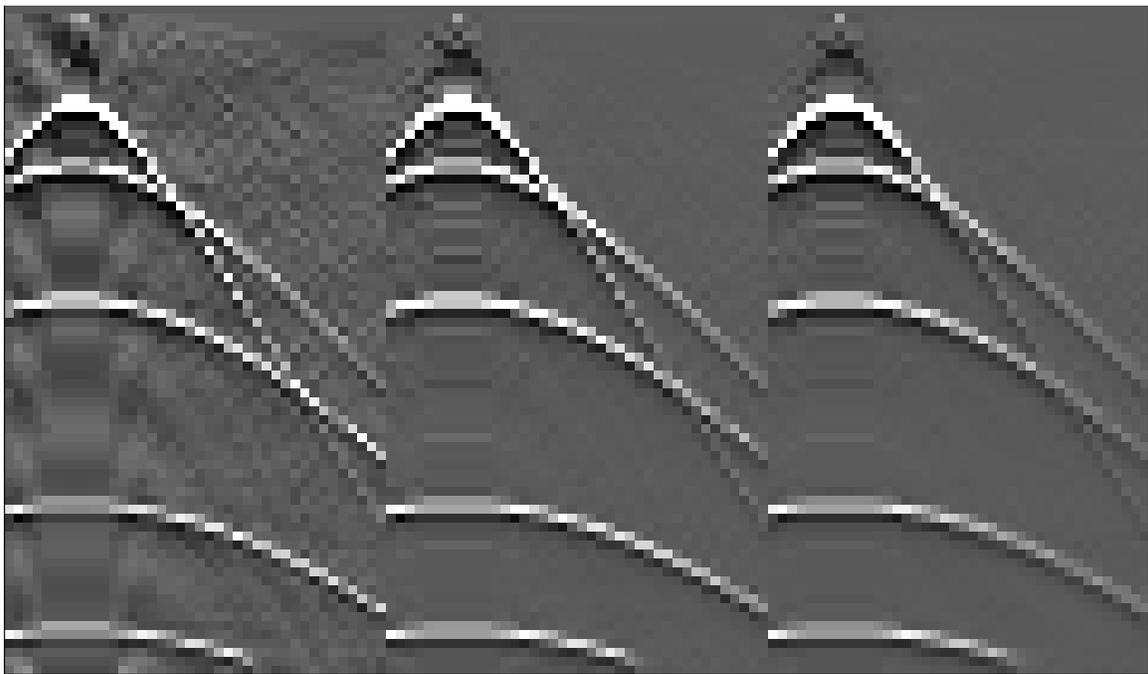


Figure 6: From left to right: 1) Remodeled data with l^2 . 2) Remodeled data with the Huber solver. 3) Remodeled data using IRLS `antoine2-comp-data-freq30` [ER]

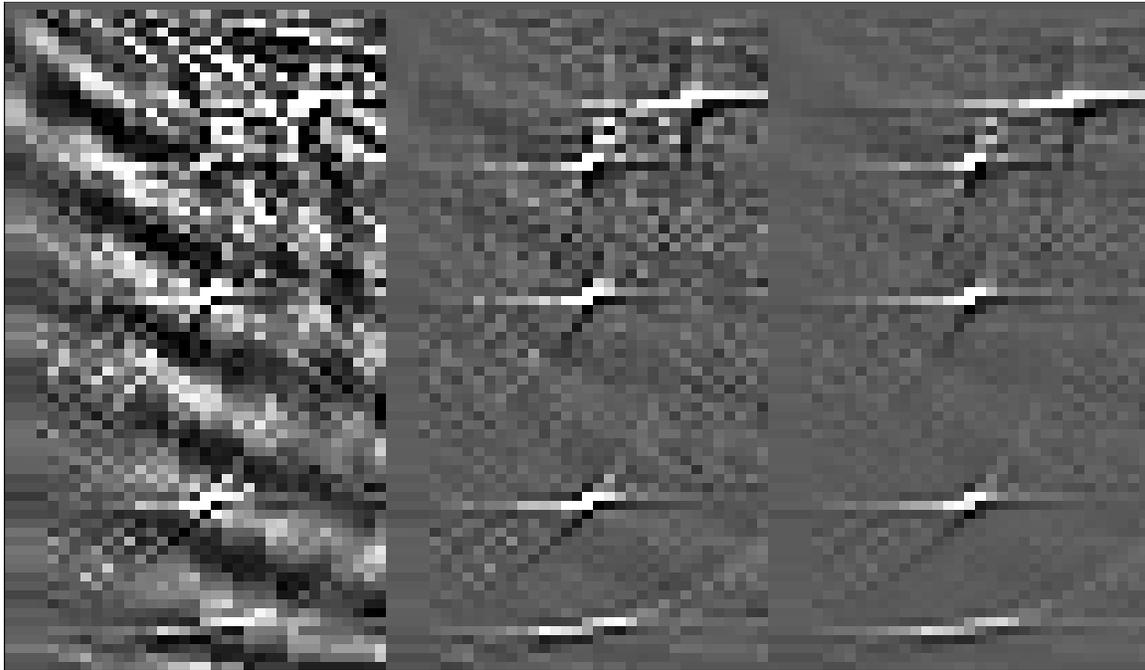


Figure 7: From left to right: 1) l^2 velocity space. 2) Huber velocity space. 3) IRLS velocity space `antoine2-comp-mod-freq30` [ER]

artifacts than they do in the l^2 case.

Convergence issues

In this section I compare the convergence of the two solvers. Because the weighting matrix \mathbf{W} is recomputed after a certain number of iterations, the IRLS algorithm consists of a series of different inverse problems. In this case, it may be unrealistic to compare the convergence of the Huber solver with IRLS. Nonetheless, because the final model should be nearly identical for both methods (assuming that they converge to same solution), the comparison appears to be fair. Note that for the IRLS algorithm, the residual is computed using equation 4. Figures 11, 12 and 13 show respectively the convergence of the two inversion schemes for the spiky data, the data with groundroll, and the real data. Except for the spiky data (Figure 11), the Huber solver converges better than the IRLS scheme. Note that small jumps in the residual of the IRLS algorithm usually appear when the weighting matrix \mathbf{W} is recomputed. Figures 14 and 15 display the convergence of the inverse problem with the spiky data for different thresholds. A thorough comparison with Figure 11 (which corresponds to the “optimal” threshold given by Darche (1989)) proves that the convergence is rather sensitive to this parameter: a bigger threshold (Figure 15) leads to better convergence, whereas a smaller value decreases it. In both cases, the Huber solver converges significantly better. Last, Figure 16 displays the convergence of the IRLS algorithm for different restart parameters: the weighting matrix is recomputed after every iteration (steepest descent), after every five and after every 15 it-

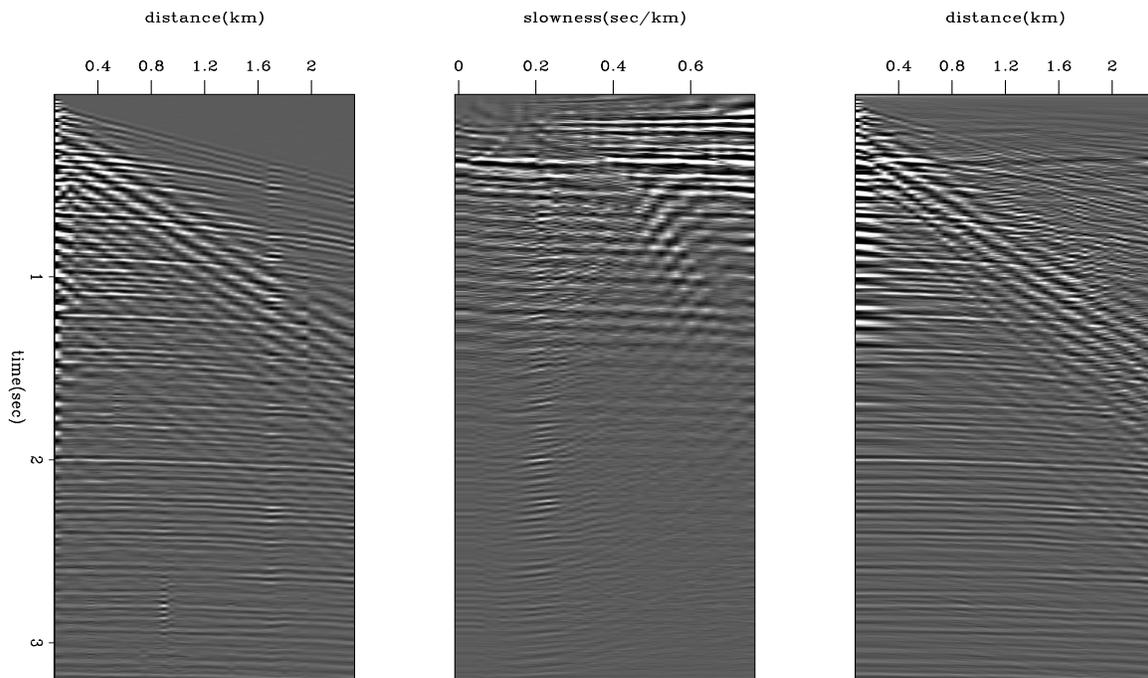


Figure 8: From left to right: 1) Input data. 2) velocity domain, l^2 inversion result. 3) remodeled data from the l^2 result [antoine2-gr-wz08-L2-HUBER](#) [ER]

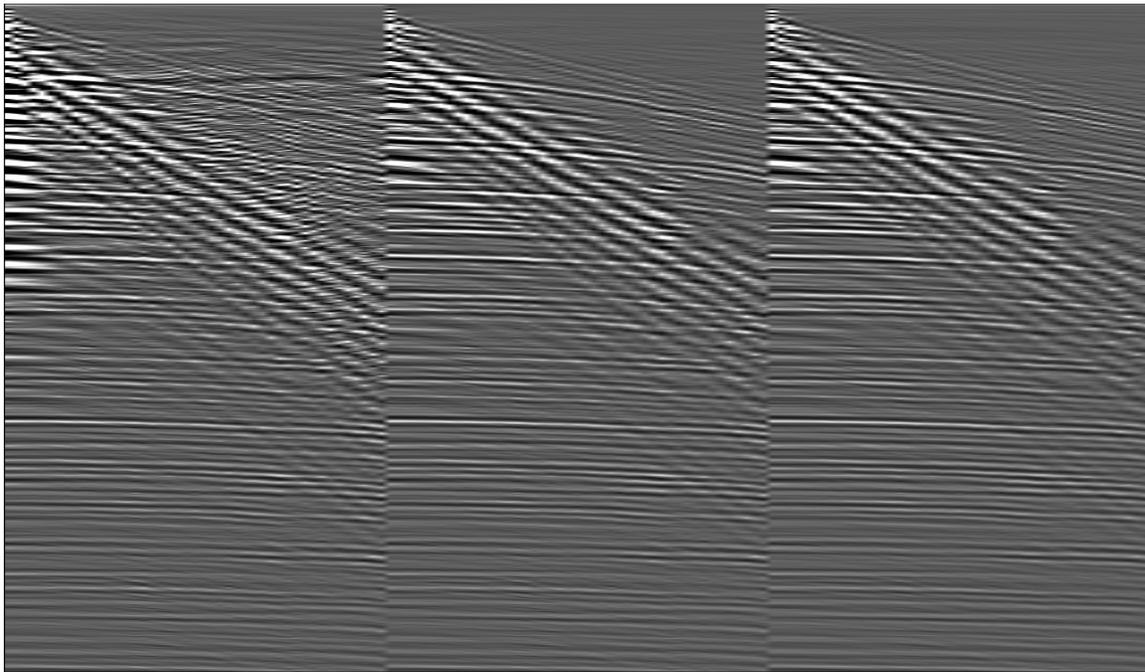


Figure 9: From left to right: 1) Remodeled data with l^2 . 2) Remodeled data with the Huber solver. 3) Remodeled data using IRLS [antoine2-comp-data-wz08-0.082](#) [ER]

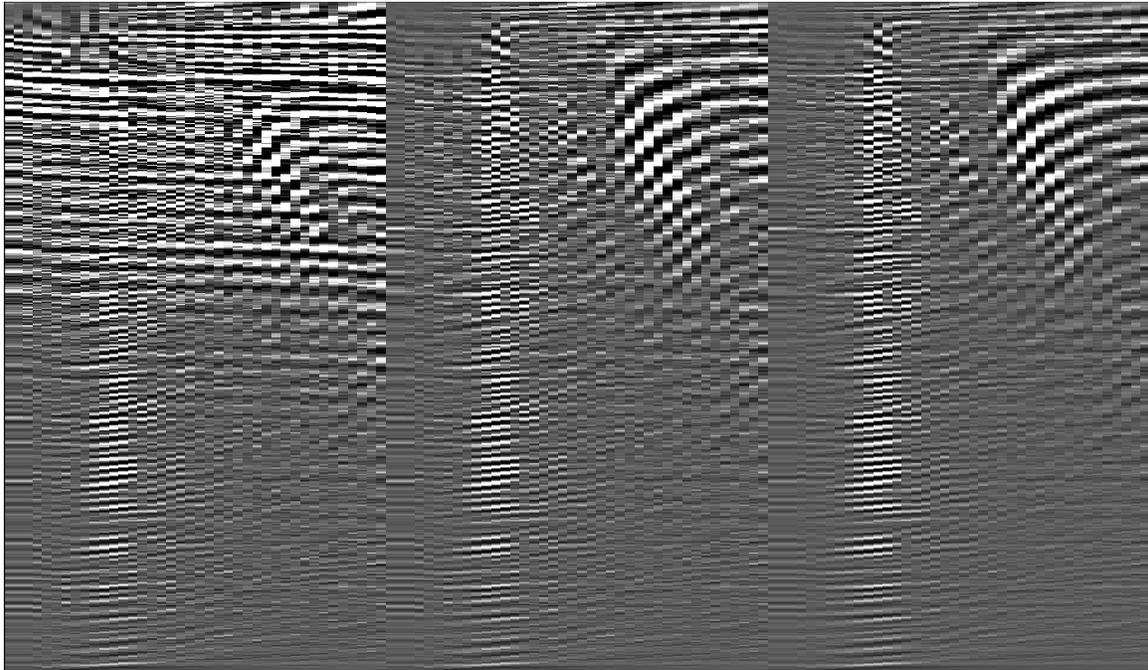


Figure 10: From left to right: 1) l^2 velocity space. 2) Huber velocity space. 3) IRLS velocity space `antoine2-comp-mod-wz08-0.082` [ER]

erations. Recomputing the weight after fifteen iterations increases the convergence, but only slightly. As expected, recomputing the weight after every iteration considerably slows down the convergence.

More about IRLS

All the convergence issues discussed above are dramatically altered if I change the restart parameter, the weighting function, and the damping factor. This means we have to make many decisions about a single inverse problem. To illustrate even more clearly the complexity of IRLS algorithms, I below give a list of weighting functions I found in the literature (ϵ positive constant):

$$w_{ii} = \frac{1}{[1 + (r_{ii}/\epsilon)^2]^{1/4}},$$

(Bube and Langan, 1997),

$$w_{ii} = \frac{2\bar{r}}{|r_{ii}| + r_{ii}},$$

(Fomel and Claerbout, 1995),

$$w_{ii} = \begin{cases} |r_{ii}|^{(p-2)/2}, & |r_{ii}| \geq \epsilon \\ |\epsilon|^{(p-2)/2}, & |r_{ii}| < \epsilon \end{cases}$$

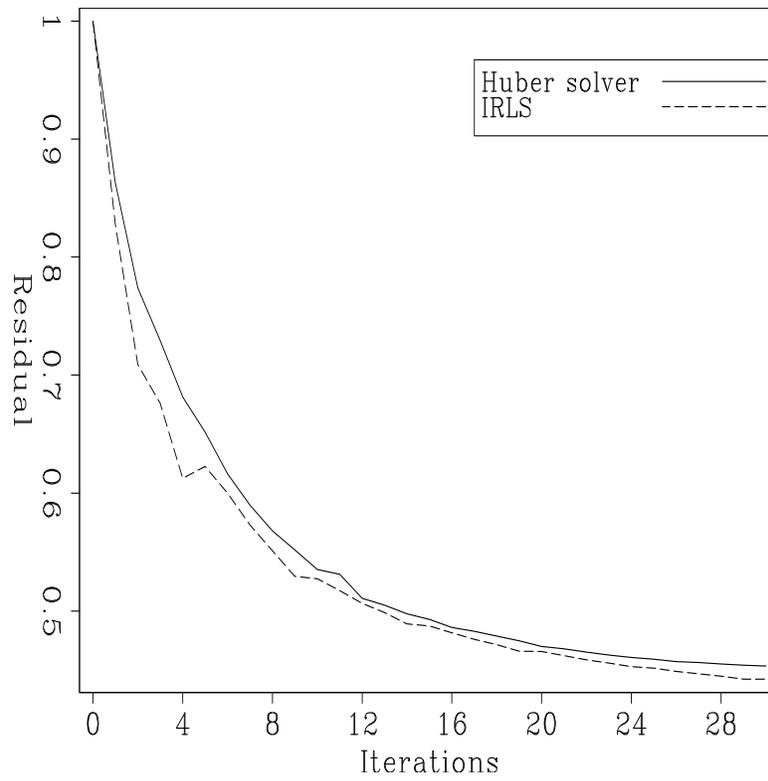


Figure 11: Relative residuals, spiky data `antoine2-r-spiky-0.04` [ER]

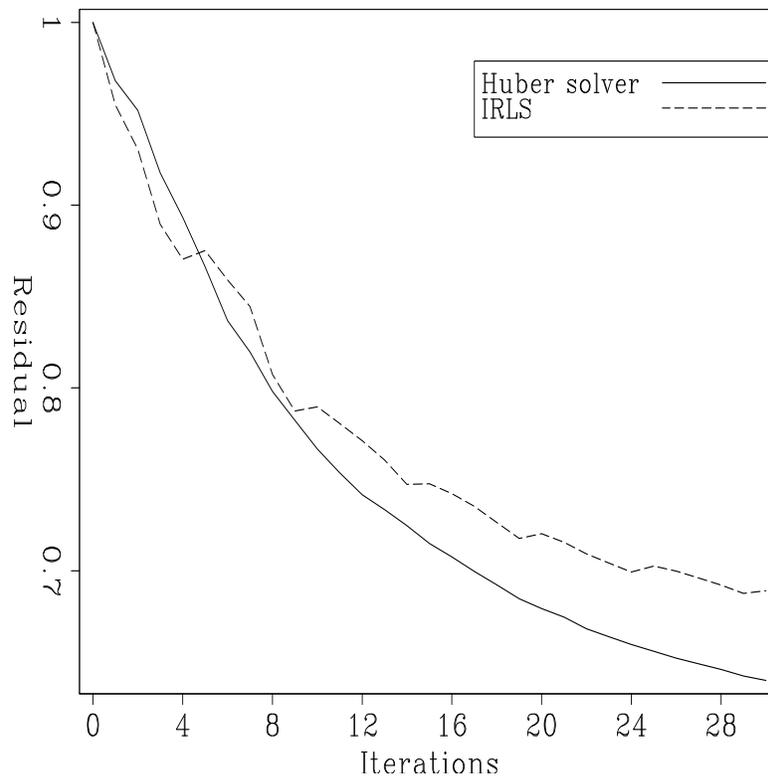


Figure 12: Relative residuals, groundroll data `antoine2-r-freq30` [ER]

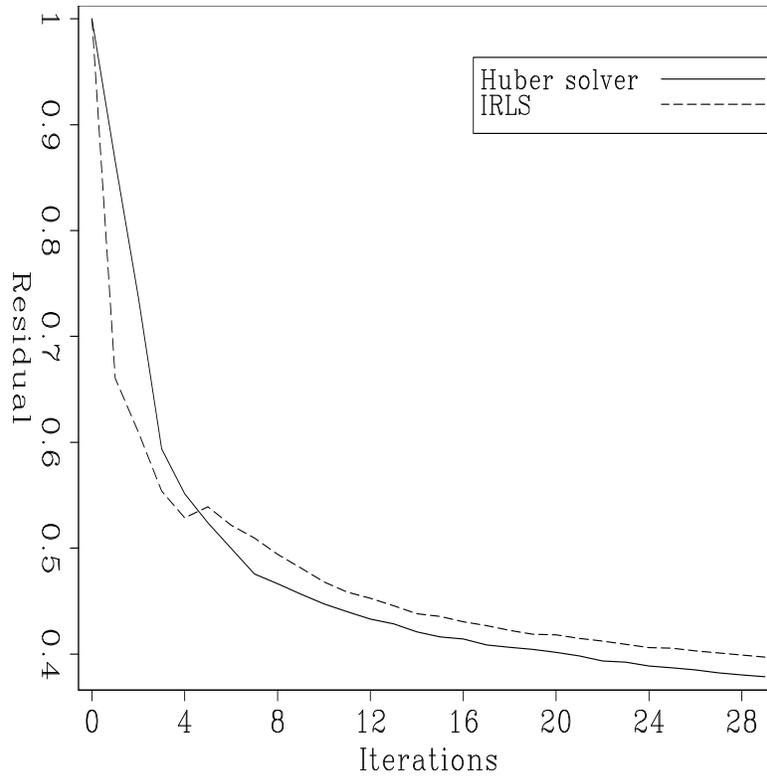


Figure 13: Relative residuals, real data `antoine2-r-wz08-0.082` [ER]

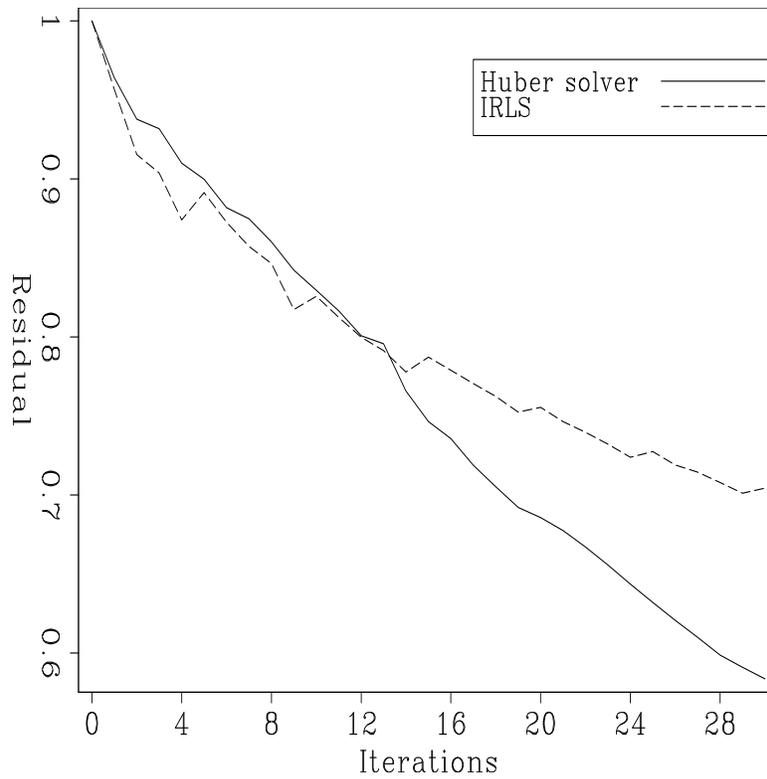


Figure 14: Relative residuals, spiky data: $\epsilon = 0.001$ `antoine2-r-spiky-0.001` [ER]

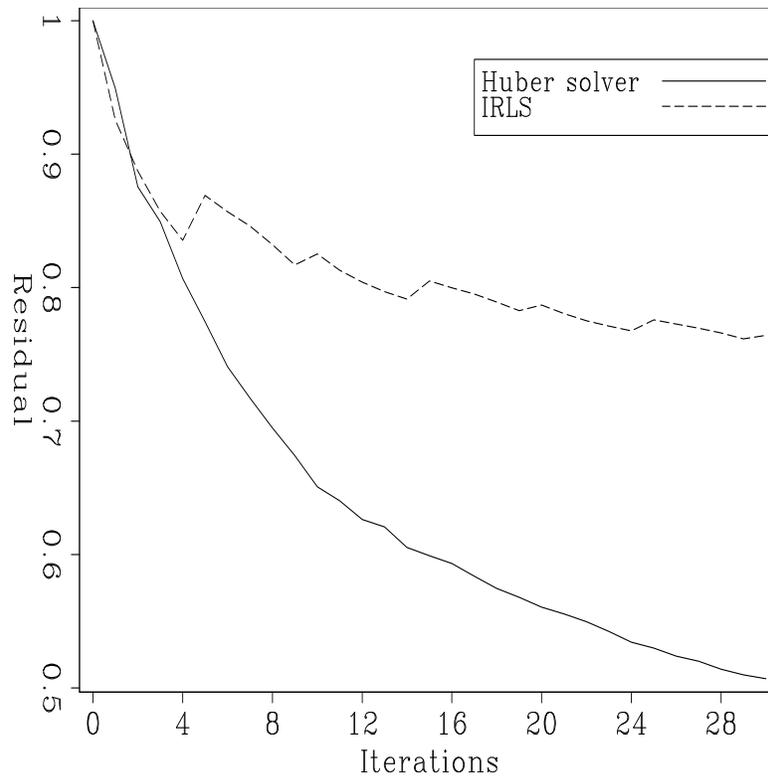


Figure 15: Relative residuals, spiky data: $\epsilon = 4$ `antoine2-r-spiky-4` [ER]

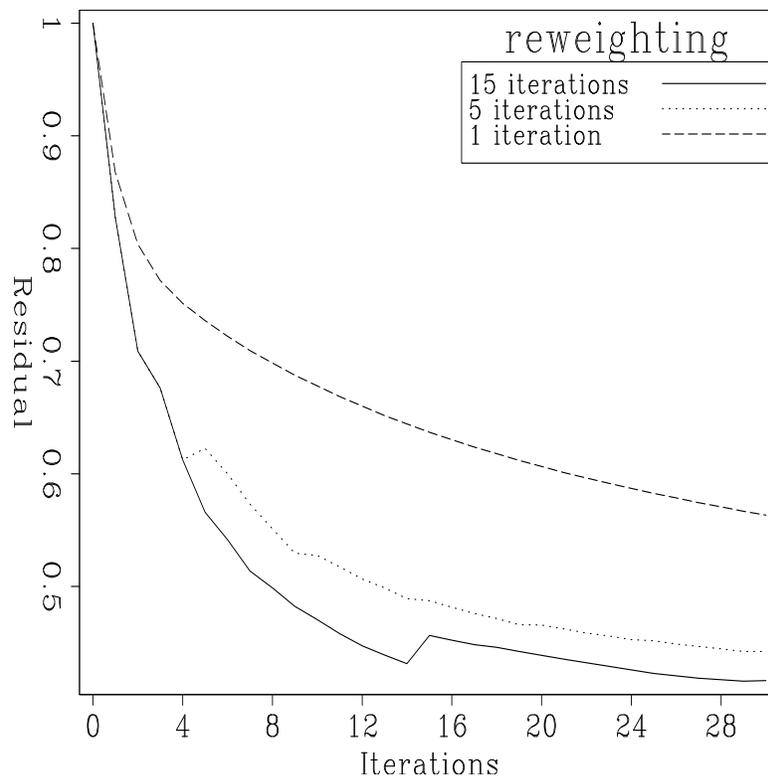


Figure 16: IRLS convergence for different restarts `antoine2-cIRLSrt` [ER]

with $p=1$, (Huber, 1981),

$$w_{ii} = \frac{p^{1/2}}{(|r_{ii}| + \epsilon)^{(2-p)/2}},$$

with $p=1$, (Hugonnet, 1998), and finally

$$w_{ii} = \frac{1}{\sqrt{1 + (r_{ii}/\epsilon)^2}}.$$

(Claerbout and Fomel, 1999), and there is probably more.

Each weighting function has different pros and cons and should be carefully chosen according to the problem we are trying to solve. Generally speaking, they all aim to weight down outliers in the data. With only the threshold to set up a priori, the Huber solver appears far easier to utilize than IRLS algorithms. Moreover, on a velocity stack inversion problem, for a very common weighting function (equation 5) with reasonable parameters (damping factor and restart parameter), I have shown that the Huber norm fosters better convergence than IRLS.

CONCLUSION

With thoroughly chosen parameters, I have shown that the l^1 problem can be approached using either the Huber along with an appropriate non-linear solver or with IRLS algorithms along with an appropriate weighting function. For the particular problem of velocity stack inversion, however, the Huber solver generally bears better convergence properties. From the practical point of view, the IRLS algorithms pose cumbersome problems for the choice of (1) the weighting function (2) the restart parameter and (3) the damping factor. With only one parameter to set up a priori, the Huber solver appears far easier to operate. This ease of use should encourage more people to use the Huber norm when implementing the hybrid l^2 - l^1 norm.

REFERENCES

- Bube, K. P., and Langan, R. T., 1997, Hybrid λ_1/λ_2 minimization with applications to tomography: *Geophysics*, **62**, no. 04, 1183–1195.
- Claerbout, J. F., and Black, J. L., 1997, Basic earth imaging: Class notes, <http://sepwww.stanford.edu/sep/prof/index.html>.
- Claerbout, J. F., and Fomel, S., 1999, Geophysical estimation with example: Class notes, <http://sepwww.stanford.edu/sep/prof/index.html>.
- Darche, G., 1989, Iterative 11 deconvolution: SEP-Report, **61**, 281–301.
- Fomel, S., and Claerbout, J., 1995, Searching the sea of Galilee: SEP-**84**, 259–270.

- Guitton, A., and Symes, W. W., 1999, Robust and stable velocity analysis using the Huber function: 69th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1166–1169.
- Guitton, A., 2000, Implementation of a nonlinear solver for minimizing the Huber norm: SEP-103, 281–289.
- Huber, P. J., 1973, Robust regression: Asymptotics, conjectures, and Monte Carlo: Ann. Statist., **1**, 799–821.
- Huber, P. J., 1981, Robust statistics: Wiley series in Probability and Mathematical statistics.
- Hugonnet, P., 1998, La transformee de radon generalisee et ses applications a la sismique: Ph.D. thesis, Ecole doctorale des sciences exactes et de leurs applications, Universite de Pau et des pays de l'Adour.
- Kelley, C. T., 1999, Iterative methods for optimization: SIAM in applied mathematics.
- Liu, D. C., and Nocedal, J., 1989, On the limited memory bfgs method for large-scale optimization: Mathematical Programming, **45**, 503–528.
- Lumley, D., Nichols, D., and Rekdal, T., 1994, Amplitude-preserved multiple suppression: SEP-82, 17–24.
- More, J. J., and Thuente, J., 1994, Line search algorithms with guaranteed sufficient decrease: ACM Transactions on Mathematical Software, **20**, 286–307.
- Nichols, D., 1994, Velocity-stack inversion using lp norms: SEP-82, 1–16.
- Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, **95**, 339–353.
- Paige, C. C., and Saunders, M. A., 1982, An algorithm for sparse linear equations and sparse least squares: ACM Transactions on Mathematical Software, **61**, 43–71.
- Scales, J. A., and Gersztenkorn, A., 1988, Robust methods in inverse theory: Inverse Problems, **4**, 1071–1091.
- Thorson, J. R., and Claerbout, J. F., 1985, Velocity-stack and slant-stack stochastic inversion: Geophysics, **50**, 2727–2741.

Speculations on contouring sparse data: Gaussian curvature

Jon Claerbout and Sergey Fomel¹

ABSTRACT

We speculate about regularizing (interpolating) sparse data. We speculate that L_1 regularization would be desirable. An example convinces us it would not. Changing direction we learn that flexed paper has zero Gaussian curvature. Unfortunately, Gaussian curvature is a nonlinear function of the altitude.

INTRODUCTION

Twenty-five years ago I² attended a series of lectures organized by the University of Houston called “Petroleum Geology for Geophysicists”. One of the professors, Daniel Busch (if I recall correctly), proposed a data set that would be “interesting to contour”. He might have said that specialized knowledge of petroleum reservoirs would be helpful. His experience was with very sparse data. I recall it being well logs from Mexico. Of special interest were (and are) sand thicknesses. He cited four wells, each with a measurement:

.	5	98
.		
.		x
.		
.	100	7

The question for the interpreter is, what should be the value of x ? Mathematical algorithms generally give a value of x near 53. As a petroleum geologist, Busch was accustomed to visualizing drainage patterns such as rivers with residual sands. For a river running northwesterly, the value of x would be near 6. On the other hand, he said, a paleotopography also commonly contains ridges, so maybe x should be roughly 100.

This example charmed me enough to remember it for 25 years and to relay it to you now, with the hope that we can do something helpful about Busch’s problem (that mathematical contouring is not as good as common sense). The solutions that we are most accustomed to are the linear solutions that come from minimizing quadratic forms. Such setups generally give the average value of x near 54 that Busch would like to avoid.

¹email: claerbout@stanford.edu, sergey@sep.stanford.edu

²In this paper “I” refers to the first author alone.

EXAMPLES

Median of an even number of points

Consider four numerical values, say (5,7,98,100). The median value is the middle value. Since this is an even number of points, there is no middle. A way to define the median m mathematically is to choose m to minimize

$$q = |m - 5| + |m - 7| + |m - 98| + |m - 100| \quad (1)$$

If you plot the function $q(m)$ you find it has a flat spot between $m = 7$ and $m = 98$. This illustrates a principle of L_1 optimizations: The minimum residual is unique. The model, however, is not unique, but is a range. This is an interesting feature of L_1 which differs from our usual least squares L_2 experience where we never get intervals. With L_2 , solutions are unique, except for a possible null space, an infinite family of added solutions. Having a 5th data point, even if very weakly weighted, would resolve the ambiguity so we might be on the track of a Busch-like solution.

Best fitting straight line

Consider the straight line that best fits a collection of data points. Suppose there are four points. Two of them are $(-10, -10)$, $(-10, 10)$. I'll call these two data points the "left slot". The other two are $(10, -10)$, $(10, 10)$ which I'll call the "right slot". You can easily see that straight lines that lie within both the left slot and the right slot all have the same sum of absolute distances from the line to the data. For each slot the sum is 20 so the total is 40. Thus L_1 gives us many lines inside the slots but it does not select any particular line. (This example is said to come from Albert Tarantola.)

You might object to having two data points at the same coordinate. By moving them apart a little, we suppose the "degeneracy" is broken, that a unique line becomes defined. Perhaps so. Never the less, it is clear that the residual is "almost minimum" for all lines inside the slots, and it is much bigger for lines outside the slots. Thus the reality of the slots remains, even where technically we might avoid them. Again, L_1 has the appealing feature that an additional data point, even if weighted weakly, could break the ambiguity.

Statics

An important example is the estimation of source and receiver time corrections. Here one has a set of observed traveltimes from the i th source to the j th receiver. After known systematic geometrical and velocity effects are removed, the time residual matrix t_{ij} remains. Then, near-source traveltimes s_i and near-receiver traveltimes r_j are estimated from the t_{ij} by minimizing the error e_{ij} in

$$e_{ij} = t_{ij} - s_i - r_j. \quad (2)$$

A trivial nonuniqueness is that an arbitrary constant added to all the s_i and subtracted from all the r_j will give the same residuals. I was surprised to discover deeper nonuniqueness lurking in a simple example. Absolute error minimization reduced a 3-by-3 matrix of t_{ij} to the e_{ij} residual matrix

$$e_{ij} = \begin{bmatrix} 0 & -12 & 4 \\ 17 & 0 & 0 \\ 0 & 10 & 0 \end{bmatrix}. \quad (3)$$

As expected theoretically (by the solution method I used), there are 5 zeros representing the 5 independent unknowns of the 6 unknowns. Note that $\sum |e_{ij}| = 43$. Now modify source and receiver times by applying +12 to row 1 and -12 to column 1. We have

$$\begin{bmatrix} 0 & 0 & 16 \\ 5 & 0 & 0 \\ -12 & 10 & 0 \end{bmatrix}, \quad (4)$$

still with $\sum |e_{ij}| = 43$. Now apply +12 to row 3 and -12 to column 3. We have

$$\begin{bmatrix} 0 & 0 & 4 \\ 5 & 0 & -12 \\ 0 & 22 & 0 \end{bmatrix}. \quad (5)$$

Furthermore, we can generate an infinite set of e_{ij} (and hence source and receiver corrections) all with the same $\sum |e_{ij}|$ by taking residuals (3)-(5) and forming any convex combination (weighted combination where each weight is positive and the weights sum to one).

The existence of a sizeable nonuniqueness with absolute error minimization leaves us the uncomfortable feeling that the mathematical uniqueness of squared error is not genuine, i.e., that the uniqueness of results with squared error is not a realistic characterization of our certainty.

Often, however, the this unfamiliar nonuniqueness does not arise. It depends on the data, not the mathematical structure of the problem. For example, I don't know any other minimum L_1 solutions with the e_{ij} matrix:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 7 & -11 \\ 0 & -3 & 8 \end{bmatrix}. \quad (6)$$

More details are found in (Claerbout and Muir, 1973) which is where I recovered this example.

Curve through two points

Consider values along a horizontal x -axis ranging from 1 to 100. Suppose at $x = 1$, the y value is given to be $y_1 = 1$. Likewise at $x = 100$ the y value is given to be $y_{100} = 100$. Now we are to find all the intervening points, y_2, y_3, \dots, y_{99} . Let us use the L_1 criterion

$$\min_{y_2, y_3, \dots, y_{99}} |y_2 - y_1| + |y_3 - y_2| + |y_4 - y_3| + \dots \quad (7)$$

The solution to this problem is any curve with a positive slope since all such curves result in the same value of 99 for (7) That is quite a lot of curves!

Now we begin to appreciate the strange flavor of L_1 . We appreciate the idea that solutions are “intervals”. But it is distressing to realize that we could often have graphical difficulty displaying the results. In practice we might need to settle for “seeing some examples.” Perhaps a satisfactory way of generating those examples would be by using random starting values for the fitting.

Suppose we set up the Busch problem with L_1 . Perhaps we will find the solution is not a unique surface. It might turn out to be a “mat” of variable thickness. It would be annoying to try to display the thickness, but perhaps the thickness is related to the uncertainty of the result. That should have value.

WHAT IS THE L_1 NORM OF THE 2-D GRADIENT?

The idea of finding smooth solutions is to minimize a measure of the gradient. The first time I thought about doing this with L_1 , I tried the wrong approach (and that put me off the track for 25 years). The wrong approach is to take the L_1 norm of the x -component of the gradient and add it to the L_1 norm of the y -component of the gradient. This is bad because it embeds the orientation of the coordinate system. Axiomatically, in science we like solutions that are independent of the human choice of a coordinate system. Thus L_1 appears to conflict with this basic requirement.

An approach independent of coordinate rotation and translation on a grid is to minimize

$$q(u) = \sum \sqrt{\nabla u \cdot \nabla u} \quad (8)$$

where $u = u(x, y)$ and where the summation is over (x, y) -space. Multivariate L_1 -norm problems generally reduce to a line search that is a weighted median. Hoare’s algorithm makes this very fast. Unfortunately, this multidimensional generalization of L_1 does not seem to reduce to a weighted median so Hoare’s algorithm is irrelevant, as might be other L_1 experiences we have seen in 1 – D .

I discussed the $\sum \sqrt{\nabla u \cdot \nabla u}$ criterion for a while with Bill Symes. We came up with this simple problem where we would use zero side boundaries and seek the response of an impulse in the medium.

$$\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 0 & a & 10 & 0 \\ 0 & b & c & 0 \\ 0 & 0 & 0 & 0 \end{array}$$

The free variables are a , b , and c . We take the x derivative diagonally to the northeast and the

y derivative diagonally to the southeast.

$$q(a,b,c) = \frac{|a|}{\sqrt{b^2+a^2}} + \frac{\sqrt{10^2+a^2}}{\sqrt{(a-c)^2+(b-10)^2}} + \frac{10}{\sqrt{10^2+c^2}} + \frac{|b|}{\sqrt{b^2+c^2}} + c \quad (9)$$

A few manual calculations quickly convinced us that the solution is $a = b = c = 0$. Thus multidimensional L_1 does not look like the answer we seek. It looks like the boundaries at infinite distance dominate the data (in this case the 10). Thus, the response to an isolated collection of spikes, might simply be the spike values where they are given and zero elsewhere.

GAUSSIAN CURVATURE

I proposed that we find out the differential equation that describes the bending of paper and use it as a regularization. The idea is to encourage a Busch-like behavior. As with L_1 , I would like to have a linear operator to preserve the uniqueness of the solution. Uniqueness gives reliability. My experience has taught me that if a method has multiple isolated minima, I will descend into the wrong one. If the paper-bending operator is nonlinear, I could linearize it.

Bill Symes suggested the Gaussian curvature. My favorite search engine (google.com) quickly gave me several references. Indeed a sheet of paper does seem to have a Gaussian curvature of zero. The Gaussian curvature of a 2-D function vanishes wherever the the function is locally one dimensional. The Gaussian curvature is the product of the principal curvatures. The Gaussian curvature is

$$\frac{h_{xx}h_{yy} - h_{xy}^2}{1 + h_x^2 + h_y^2} \quad (10)$$

For small dips, the numerator is the important part. The numerator is the determinant of the Hessian,

$$\det \begin{vmatrix} h_{xx} & h_{xy} \\ h_{yx} & h_{yy} \end{vmatrix} \quad (11)$$

We might regularize a collection of data points by minimizing this determinant. I have begun looking for references that may have previously investigated this very basic idea. Unfortunately, the function is nonlinear. We can linearize it. Replacing h by $\bar{h} + h$ and dropping terms in h^2 we get

$$0 \approx (\bar{h}_{xx}\bar{h}_{yy} - \bar{h}_{xy}^2) + \bar{h}_{xx}h_{yy} + h_{xx}\bar{h}_{yy} - 2\bar{h}_{xy}h_{xy} \quad (12)$$

The most important question is: what is $\bar{h}(x,y)$? How do we initialize it, and how can we safely update it? A way to initialize $\bar{h}(x,y)$ is to approximate the initial data by a best fitting one-dimensional parabola. One way to stabilize $\bar{h}(x,y)$ is to smooth it in patches.

I am reminded of "LOMOPLAN", an earlier idea I had to fit a best plane wave, then use it to define a linear operator to use as a weighting function in estimation. The idea is that a sedimentary section consists of a single local plane wave. Perhaps that two-stage least squares process is akin to linearizing the Gaussian curvature.

MORE ON GAUSSIAN CURVATURE

Given a function $u(x, y)$, its x -derivative u_x , its y -derivative u_y , and a slope parameter p , we have the planewave operator L

$$0 = L(u) = u_x + pu_y \quad (13)$$

which vanishes when $u(x, y)$ is not really a two-dimensional function but is a one-dimensional function $u = f(x - py)$.

Next we get two equations from the plane-wave equation, one differentiating by x , the other by y .

$$0 = L(u_x) = u_{xx} + pu_{xy} \quad (14)$$

$$0 = L(u_y) = u_{yx} + pu_{yy} \quad (15)$$

Eliminate p from these two equations by solving for it:

$$-p = \frac{u_{xx}}{u_{xy}} = \frac{u_{yx}}{u_{yy}} \quad (16)$$

or

$$G(u) = u_{xx}u_{yy} - u_{xy}u_{yx} = 0 \quad (17)$$

The plane-wave operator $L(u)$ will not vanish unless u is a plane wave going in the direction specified by p . A remarkable property of the function $G(u)$ is that it vanishes for any orientation of plane wave. If we want to test a 2-D field for one-dimensionality, the test $L(u)$ requires us to know p . The test $G(u)$ does not. Generally we do not know p and we need to estimate it by statistical means in a window of some size that we must specify. In principle, G escapes those problems (although it might be worse in practice because it is a nonlinear function of the wavefield).

In differential geometry, a quantity appears that is known as the ‘‘Gaussian curvature’’. For small vertical motions u , this Gaussian curvature reduces to our expression G .

The gentle flexure of a sheet of paper follows the principle that the Gaussian curvature vanishes. The deformation must be one dimensional. We were first attracted to Gaussian curvature as a way of interpolating sparse data where the data represents a wavefield or a sedimentary earth model. We would seek the interpolation that was most ‘‘paper like’’, which minimized the integral of the square of the Gaussian curvature. Unfortunately, G is already a quadratic function of u even before we square G to minimize a positive value.

Thin-plate versus biharmonic equation

The biharmonic equation uses the Laplacian operator twice: The biharmonic equation results from minimizing the quadratic form

$$B(u) = u'(\partial_{xx} + \partial_{yy})'(\partial_{xx} + \partial_{yy})u \quad (18)$$

$$B(u) = u'(\partial'_{xx}\partial_{xx} + \partial'_{xx}\partial_{yy} + \partial'_{yy}\partial_{xx} + \partial'_{yy}\partial_{yy})u \quad (19)$$

To minimize, simply cancel off u' and set to zero. The thin plate equation resembles the biharmonic equation but differs in a subtle but important way. The quadratic form minimized for a thin plate is:

$$T(u) = v'v \quad \text{where } v = \begin{bmatrix} \partial_{xx} \\ \partial_{xy} \\ \partial_{yx} \\ \partial_{yy} \end{bmatrix} u \quad (20)$$

or

$$T(u) = u'(\partial'_{xx} \partial_{xx} + \partial'_{yy} \partial_{yy} + \partial'_{xy} \partial_{xy} + \partial'_{yx} \partial_{yx})u \quad (21)$$

Again, we find the associated differential equation by canceling off the u' .

What is bothering me is that the dispersion relations look the same but the quadratic forms look different. The difference between the biharmonic quadratic form and the thin plate quadratic form lies in the cross term. Let us form half this difference $G = (B - T)/2$.

$$G(u) = u_{xx}u_{yy} - u_{xy}u_{yx} \quad (22)$$

We see the difference has turned out to be the Gaussian curvature. Although the difference is a quadratic form, it is not uniformly positive or negative, as it can have both signs.

By means of rotation of coordinates, we can diagonalize the matrix

$$\begin{bmatrix} u_{xx} & u_{xy} \\ u_{yx} & u_{yy} \end{bmatrix} \xrightarrow{\text{rotation}} \begin{bmatrix} K_1 & 0 \\ 0 & K_2 \end{bmatrix} \quad (23)$$

Thus we can think of $G = K_1K_2$ as the product of the curvatures while the biharmonic quadratic form is the square of the sum $B = (K_1 + K_2)^2$. In terms of curvatures, in the rotated coordinates the thin plate operator is

$$T = B - 2G \quad (24)$$

$$T = (K_1 + K_2)^2 - 2K_1K_2 \quad (25)$$

$$T = K_1^2 + K_2^2 \quad (26)$$

which is the sum of the squares of the curvatures.

The meaning is this: The biharmonic equation zeroes $B = (K_1 + K_2)^2$ so its solution could be expected to have many places of $K_1 = -K_2$ where the curvature on one axis is the negative of that on the other axis. In other words, solving the biharmonic equation might give us a function containing many saddles. On the other hand, the thin-plate equation $T = K_1^2 + K_2^2$ tries to eliminate both curvatures (not allowing credit for when one cancels the other). However, with respect to optimization, both quadratic forms are equivalent.

CONCLUSION

In conclusion, we do not see any immediate action items. The notion of minimizing Gaussian curvature is appealing, but it is nonlinear, which means that solutions depend on the starting location. Physically, when flexing paper, the final deformation probably depends on the deformation history. For practical purposes the thin plate operator is the Laplacian squared. If we are going to try to minimize the Gaussian curvature, a nonlinear criteria, we should probably begin from the thin plate which is unique.

REFERENCES

Claerbout, J. F., and Muir, F., 1973, Robust modeling with erratic data: *Geophysics*, **38**, no. 5, 826–844.

Short Note

Implementation of a nonlinear solver for minimizing the Huber norm

Antoine Guitton¹

INTRODUCTION

The Huber norm (Huber, 1973) is an alternative to Iteratively Reweighted Least Square programs for solving the hybrid l^2 - l^1 problem. In this note, I detail a method for minimizing the Huber norm. Because the Huber norm gives rise to a non-linear problem with non-twice continuously differentiable objective functions, its use is quite challenging. Claerbout (1996) implemented a Huber regression based on conjugate-gradient descents. However, the final results were not satisfying. Here I propose to solve the Huber problem using a quasi-Newton update of the solution with the computation of an approximated Hessian (second derivative of the objective function). This strategy is innovative in seismic processing and merits some explanation.

In this paper I first provide general definitions plus sufficient conditions to solve the optimization problem. Then, I present the quasi-Newton method and the complete algorithm used to solve the Huber problem.

DEFINITIONS AND CONDITIONS FOR OPTIMALITY

This part follows closely Kelley's *Iterative Method for Optimization* (Kelley, 1999). We start here with a series of definitions:

1. \mathbf{A} is positive definite if $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ for all $\mathbf{x} \in \mathfrak{R}^N$
2. \mathbf{A} is *spd* if \mathbf{A} is positive definite and symmetric
3. $\mathbf{x}^* \in U$ ($U \subset \mathfrak{R}^N$) is a global minimizer if $f(\mathbf{x}^*) \leq f(\mathbf{x})$ for all $\mathbf{x} \in U$

The Euclidian norm is also defined as

$$\|\mathbf{x}\| = \sqrt{\sum_{i=1}^N (x_i)^2}.$$

¹email: antoine@sep.stanford.edu

Now, I give sufficient conditions that a minimizer \mathbf{x}^* exists for a function f .

Theorem

Let f be twice continuously differentiable in a neighborhood of \mathbf{x}^* . Assume that $\nabla f(\mathbf{x}^*) = 0$ and that $\nabla^2 f(\mathbf{x}^*)$ is positive definite, then \mathbf{x}^* is a local minimizer of f .

Proof

Let $\mathbf{u} \in \mathfrak{R}^N$ with $\mathbf{u} \neq 0$. For sufficiently small t we have

$$f(\mathbf{x}^* + t\mathbf{u}) = f(\mathbf{x}^*) + t\nabla f(\mathbf{x}^*)^T \mathbf{u} + \frac{t^2}{2} \mathbf{u}^T \nabla^2 f(\mathbf{x}^*) \mathbf{u} + o(t^2).$$

But $\nabla f(\mathbf{x}^*) = 0$ giving

$$f(\mathbf{x}^* + t\mathbf{u}) = f(\mathbf{x}^*) + \frac{t^2}{2} \mathbf{u}^T \nabla^2 f(\mathbf{x}^*) \mathbf{u} + o(t^2).$$

If $\nabla^2 f(\mathbf{x}^*)$ is positive definite, its smallest eigenvalue λ obeys $\lambda > 0$. So we have

$$f(\mathbf{x}^* + t\mathbf{u}) - f(\mathbf{x}^*) \geq \frac{\lambda}{2} \|t\mathbf{u}\|^2 + o(t^2) > 0.$$

Then, \mathbf{x}^* is a local minimizer for f .

We see that a sufficient condition for a local minimizer is $\nabla f(\mathbf{x}^*) = 0$ and $\nabla^2 f(\mathbf{x}^*)$ (Hessian) is positive definite. These conditions are very important and should guide us in the choice of an optimization strategy.

Quadratic functions form the basis for most of the algorithms in optimization, in particular for the quasi-Newton method detailed in this paper. It is then important to discuss some issues involved with these functions. Now, if we pose a quadratic objective function

$$f(\mathbf{x}) = -\mathbf{x}^T \mathbf{b} + \frac{1}{2} \mathbf{x}^T \mathbf{H} \mathbf{x},$$

we see that we want to solve

$$\nabla f(\mathbf{x}) = -\mathbf{b} + \mathbf{H} \mathbf{x} = 0.$$

We may assume that the Hessian \mathbf{H} is symmetric because

$$\mathbf{x}^T \mathbf{H} \mathbf{x} = \mathbf{x}^T \frac{\mathbf{H}^T + \mathbf{H}}{2} \mathbf{x}.$$

So, the unique global minimizer is the solution of the system above if \mathbf{H} (the Hessian) is *spd*.

A QUASI-NEWTON METHOD FOR UNCONSTRAINED OPTIMIZATION

We will assume that f and \mathbf{x}^* satisfy the following assumptions:

1. f is twice continuously differentiable
2. $\nabla f(\mathbf{x}^*) = 0$
3. $\nabla^2 f(\mathbf{x}^*)$ is symmetric positive definite

The Newton methods update the current iteration \mathbf{x}_n by the formula

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \lambda_n \mathbf{H}_n^{-1} \nabla f(\mathbf{x}_n), \quad (1)$$

where λ_n is given by a line search that ensures sufficient decrease. Quasi-Newton methods update an approximation of the Hessian \mathbf{H}_n^{-1} as the iterations progress. A possible update is the BFGS method (Broyden, 1969; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970), which overcomes some limitations of the earlier Broyden's method (Broyden, 1965). In particular, the Broyden's update does not keep the *spd* structure of the Hessian. This structure not only ensures the existence of a local minimizer but also allows the convergence of the updated solution \mathbf{x}_{n+1} to the minimum (Kelley, 1999). The BFGS update is a rank-two update given by

$$\mathbf{H}_{n+1} = \mathbf{H}_n + \frac{\mathbf{y}\mathbf{y}^T}{\mathbf{y}^T \mathbf{s}} - \frac{(\mathbf{H}_n \mathbf{s})(\mathbf{H}_n \mathbf{s})^T}{\mathbf{s}^T \mathbf{H}_n \mathbf{s}}, \quad (2)$$

with $\mathbf{s} = \mathbf{x}_{n+1} - \mathbf{x}_n$ and $\mathbf{y} = \nabla f(\mathbf{x}_{n+1}) - \nabla f(\mathbf{x}_n)$. In practice, it is very useful to express the previous equation in terms of the inverse matrices. We then have

$$\mathbf{H}_{n+1}^{-1} = \left(\mathbf{I} - \frac{\mathbf{s}\mathbf{y}^T}{\mathbf{y}^T \mathbf{s}} \right) \mathbf{H}_n^{-1} \left(\mathbf{I} - \frac{\mathbf{y}\mathbf{s}^T}{\mathbf{y}^T \mathbf{s}} \right) + \frac{\mathbf{s}\mathbf{s}^T}{\mathbf{y}^T \mathbf{s}}. \quad (3)$$

Lemma

Let \mathbf{H}_n be *spd*, $\mathbf{y}^T \mathbf{s} > 0$, and \mathbf{H}_{n+1} given in equation (2). Then \mathbf{H}_{n+1} is *spd*.

Proof

Starting from equation (2), we can write for all $\mathbf{z} \neq 0$ and $\mathbf{y}^T \mathbf{s} > 0$,

$$\mathbf{z}^T \mathbf{H}_{n+1} \mathbf{z} = \mathbf{z}^T \mathbf{H}_n \mathbf{z} + \frac{(\mathbf{z}^T \mathbf{y})^2}{\mathbf{y}^T \mathbf{s}} - \frac{(\mathbf{z}^T \mathbf{H}_n \mathbf{s})^2}{(\mathbf{s}^T \mathbf{H}_n \mathbf{s})}.$$

Since \mathbf{H}_n is *spd*, we have

$$(\mathbf{z}^T \mathbf{H}_n \mathbf{s})^2 \leq (\mathbf{s}^T \mathbf{H}_n \mathbf{s})(\mathbf{z}^T \mathbf{H}_n \mathbf{z})$$

with equality only if $\mathbf{z} = 0$ or $\mathbf{s} = 0$. But we have $\mathbf{z} \neq 0$ and $\mathbf{y}^T \mathbf{s} > 0$ so that

$$\mathbf{z}^T \mathbf{H}_{n+1} \mathbf{z} > \frac{(\mathbf{z}^T \mathbf{y})^2}{\mathbf{y}^T \mathbf{s}} \geq 0.$$

Then \mathbf{H}_{n+1} is *spd*. If during the iterations we have $\mathbf{y}^T \mathbf{s} \leq 0$, then the update is a failure.

The previous lemma is very important since it shows that starting from an initial *spd* Hessian \mathbf{H}_0 , the next approximation of the Hessian is *spd* (given that $\mathbf{y}^T \mathbf{s} > 0$). This guarantees the existence of a minimizer for the function f (the inverse \mathbf{H}^{-1} is also *spd*). It can be shown (Kelley, 1999) that given some assumptions, the BFGS iterates are defined and converge *q-superlinearly*² to the local minimizer \mathbf{x}^* . In practice, the storage needed to compute the update and the possibility that $\mathbf{y}^T \mathbf{s} \leq 0$ are important issues. The updated Hessian is computed at each iteration recursively. For this, we need to store a solution step vector \mathbf{s} and a gradient step vector \mathbf{y} after each iteration. If for small problems this storage is not an issue, it may become critical for large-scale problems. Unfortunately, these large-scale problems occur in geophysics, and we need to find a better storage solution. Nocedal (1980) gives an interesting answer to this problem. Instead of keeping all the \mathbf{s} and \mathbf{y} from the past iterations, we update the Hessian using the information from the m previous iterations, where m is given by the end user. This means that when the number of iterations is smaller than m , we have a “real” BFGS update, and when it is larger than m , we have a Limited-memory BFGS (L-BFGS) update.

L-BFGS update

For the sake of completeness, I give the updating formulas of the Hessian as presented by Nocedal (1980). We define first

$$\rho_i = 1/\mathbf{y}_i^T \mathbf{s}_i \text{ and } \mathbf{v}_i = (I - \rho_i \mathbf{y}_i \mathbf{s}_i^T).$$

In addition, we pose $\mathbf{H}^{-1} = \mathbf{B}$. As described above, when k , the number of iterations, obeys $k + 1 \leq m$, where m is the storage limit, we have the usual BFGS update

$$\begin{aligned} \mathbf{B}_{k+1} &= \mathbf{v}_k^T \mathbf{v}_{k-1}^T \cdots \mathbf{v}_0^T \mathbf{B}_0 \mathbf{v}_0 \cdots \mathbf{v}_{k-1} \mathbf{v}_k \\ &\quad + \mathbf{v}_k^T \cdots \mathbf{v}_1^T \rho_0 \mathbf{s}_0 \mathbf{s}_0^T \mathbf{v}_1 \cdots \mathbf{v}_k \\ &\quad \cdot \\ &\quad \cdot \\ &\quad \cdot \\ &\quad + \mathbf{v}_k^T \rho_{k-1} \mathbf{s}_{k-1} \mathbf{s}_{k-1}^T \mathbf{v}_k \\ &\quad + \rho_k \mathbf{s}_k \mathbf{s}_k^T. \end{aligned} \tag{4}$$

² $\mathbf{x}_n \rightarrow \mathbf{x}^*$ *q-superlinearly* if

$$\lim_{n \rightarrow \infty} \frac{\|\mathbf{x}_{n+1} - \mathbf{x}^*\|}{\|\mathbf{x}_n - \mathbf{x}^*\|} = 0.$$

For $k + 1 > m$ we have the special limited-memory update

$$\begin{aligned}
\mathbf{B}_{k+1} = & \mathbf{v}_k^T \mathbf{v}_{k-1}^T \cdots \mathbf{v}_{k-m+1}^T \mathbf{B}_0 \mathbf{v}_{k-m+1} \cdots \mathbf{v}_{k-1} \mathbf{v}_k \\
& + \mathbf{v}_k^T \cdots \mathbf{v}_{k-m+2}^T \rho_{k-m+1} \mathbf{s}_{k-m+1} \mathbf{s}_{k-m+1}^T \mathbf{v}_{k-m+2} \cdots \mathbf{v}_k \\
& \cdot \\
& \cdot \\
& \cdot \\
& + \mathbf{v}_k^T \rho_{k-1} \mathbf{s}_{k-1} \mathbf{s}_{k-1}^T \mathbf{v}_k \\
& + \rho_k \mathbf{s}_k \mathbf{s}_k^T.
\end{aligned} \tag{5}$$

It is easy to show that the special updated Hessian is also *spd*. The L-BFGS algorithm is then

Algorithm 1

1. Choose \mathbf{x}_0 , m , $0 < \mu < 1$, $\mu < \nu < 1$ and a symmetric positive definite \mathbf{B}_0 . Set $k = 0$
2. Compute

$$\mathbf{d}_k = -\mathbf{B}_k \nabla f(\mathbf{x}_k) \tag{6}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{d}_k, \tag{7}$$

where λ_k verifies the Wolfe conditions (More and Thuente, 1994):

$$f(\mathbf{x}_k + \lambda_k \mathbf{d}_k) \leq f(\mathbf{x}_k) + \mu \lambda_k \nabla f(\mathbf{x}_k)^T \mathbf{d}_k, \tag{8}$$

$$|\nabla f(\mathbf{x}_k + \lambda_k \mathbf{d}_k)^T \mathbf{d}_k| \geq \nu |\nabla f(\mathbf{x}_k)^T \mathbf{d}_k|. \tag{9}$$

We always try steplength $\lambda_k = 1$ first.

3. Let $\hat{m} = \min\{k, m - 1\}$. Check if $\mathbf{y}_k^T \mathbf{s}_k > 0$.

- If no: $\mathbf{B}_{k+1} = \mathbf{I}$ (steepest descent step) and delete the pairs $\{\mathbf{y}_i, \mathbf{s}_i\}_{j=k-\hat{m}}^k$.
- If yes: Update \mathbf{B}_0 $\hat{m} + 1$ times using the pairs $\{\mathbf{y}_i, \mathbf{s}_i\}_{j=k-\hat{m}}^k$, i.e., let

$$\begin{aligned}
\mathbf{B}_{k+1} = & \mathbf{v}_k^T \mathbf{v}_{k-1}^T \cdots \mathbf{v}_{k-\hat{m}}^T \mathbf{B}_0 \mathbf{v}_{k-\hat{m}} \cdots \mathbf{v}_{k-1} \mathbf{v}_k \\
& + \mathbf{v}_k^T \cdots \mathbf{v}_{k-\hat{m}+1}^T \rho_{k-\hat{m}} \mathbf{s}_{k-\hat{m}} \mathbf{s}_{k-\hat{m}}^T \mathbf{v}_{k-\hat{m}+1} \cdots \mathbf{v}_k \\
& \cdot \\
& \cdot \\
& \cdot \\
& + \mathbf{v}_k^T \rho_{k-1} \mathbf{s}_{k-1} \mathbf{s}_{k-1}^T \mathbf{v}_k \\
& + \rho_k \mathbf{s}_k \mathbf{s}_k^T.
\end{aligned} \tag{10}$$

4. Set $k := k + 1$ and go to 2.

The update \mathbf{B}_{k+1} is not formed explicitly; instead, we compute $\mathbf{d}_k = -\mathbf{B}_k \nabla f(\mathbf{x}_k)$ with an iterative formula (Nocedal, 1980). Liu and Nocedal (1989) propose that we scale the initial symmetric positive definite \mathbf{B}_0 at each iteration:

$$\mathbf{B}_k^0 = \frac{\mathbf{y}_k^T \mathbf{s}_k}{\|\mathbf{y}_k\|^2} \mathbf{B}_0. \quad (11)$$

This scaling greatly improves the performances of the method. Liu and Nocedal (1989) show that the storage limit for large-scale problems has little effect on the method's performances. A common choice for m is $m = 5$ (this is the default in my implementation as well). Conditions (8) and (9) are satisfied if we use an appropriate line search algorithm. I programmed a MoreThuente line search algorithm (More and Thuente, 1994), which ensures sufficient decrease of the objective function (equation 8) and obeys the curvature condition given in equation (9). We do not describe this program here. In practice, the initial guess \mathbf{B}_0 for the Hessian can be the identity matrix \mathbf{I} ; then it might be scaled as proposed above. Liu and Nocedal (1989) prove that the L-BFGS algorithm converges to the local minimizer \mathbf{x}^* and that the family of solutions $\{\mathbf{x}_k\}$ converges *R-linearly*³ (remember that the usual BFGS gives *q-superlinear* convergence, which is better).

SOLVING THE HUBER PROBLEM WITH A QUASI-NEWTON METHOD

The Huber norm (Huber, 1973, 1981) is a hybrid l^1 - l^2 measure. We expect to find the minimum of the function using a quasi-Newton method with a L-BFGS update of the Hessian (Guitton and Symes, 1999). The Huber norm is

$$\begin{aligned} f(\mathbf{x}) &= |\mathbf{A}\mathbf{x} - \mathbf{m}|_{Huber}, \\ &= |\mathbf{r}|_{Huber}, \\ &= \sum_{i=1}^N M_\epsilon(r_i), \end{aligned} \quad (12)$$

where

$$M_\epsilon(r) = \begin{cases} \frac{r^2}{2\epsilon}, & 0 \leq |r| \leq \epsilon \\ |r| - \frac{\epsilon}{2}, & \epsilon < |r|. \end{cases} \quad (13)$$

ϵ commands the limit between an l^1 or l^2 treatment of the residual; we call it the Huber threshold and it must be given by the user. The gradient of the objective function is given by

$$\nabla f(\mathbf{x}) = \mathbf{A}^T (\mathbf{A}\mathbf{x} - \mathbf{m})_{-\epsilon}^\epsilon, \quad (14)$$

³ $\mathbf{x}_n \rightarrow \mathbf{x}^*$ R-linearly if there is a constant $0 \leq r < 1$ such that

$$f(\mathbf{x}_k) - f(\mathbf{x}^*) \leq r^k [f(\mathbf{x}_0) - f(\mathbf{x}^*)].$$

where $\mathbf{z}_{-\epsilon}^{\epsilon}$ is the vector whose i th component is

$$z_i = \max\{-\epsilon, \min\{\epsilon, z_i\}\}.$$

Because the Huber function is not twice continuously differentiable, it does not satisfy the three necessary conditions that guarantee the convergence to a minimum. However, we only need to compute the gradient for the BFGS update of the Hessian. Furthermore, given that the approximated Hessian is certainly a vague approximation of the real one (Symes, 1999, Personal communication), the violation of the initial conditions is mild. In addition, results (Guitton, 2000) show that this method converges to a minimum. Li (1995) shows that the Huber function has a unique minimizer for any meaningful choice of ϵ . Indeed, if the l^1 problem $f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{m}\|_1$ has multiple solutions (Tarantola, 1987), then the Huber problem, provided that ϵ is small enough, also has multiple solutions. This is annoying since we want to find a global minimum for the problem using quasi-Newton updates. In practice, however, it seems that

$$\epsilon = \frac{\max|\mathbf{d}|}{100}$$

is a good choice for the threshold (Darche, 1989). The threshold being set properly, the Huber function has mathematical properties that allow the use of quasi-Newton methods. We can now define an efficient algorithm in order to solve the Huber problem:

Algorithm 2

1. Choose \mathbf{x}_0 and the threshold ϵ . Set $k = 0$
2. Compute $\nabla f(\mathbf{x}_k)$ using equation 14
3. Compute $\mathbf{d}_k = -\mathbf{B}_k \nabla f(\mathbf{x}_k)$ using a L-BFGS update (Algorithm 1, step 3)
4. Compute the step λ_k using a MoreThuente line search ($\lambda_k = 1$ tried first)
5. Update the solution $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{d}_k$
6. Go to step 2

This algorithm will converge to the minimizer \mathbf{x}^* , as proven by Liu and Nocedal (1989).

CONCLUSION

Given an adequate threshold ϵ , the Huber problem may be solved using a quasi-Newton solver. The Limited memory BFGS method, a quasi-Newton update, has interesting storage properties that lead to efficient convergence to the local minimum of any convex function. In this paper, I proposed an algorithm to solve the Huber problem using the L-BFGS solver and a MoreThuente line search. This algorithm is then supposed to give a *R-linear* convergence to the desired solution.

REFERENCES

- Broyden, C. G., 1965, A class of methods for solving nonlinear simultaneous equations: *Math. Comp.*, **19**, 577–593.
- Broyden, C. G., 1969, A new double-rank minimization algorithm: *AMS Notices*, **16**, 670.
- Claerbout, J., 1996, Conjugate-direction Huber regression: *SEP-92*, 229–235.
- Darche, G., 1989, Iterative 11 deconvolution: *SEP-61*, 281–301.
- Fletcher, R., 1970, A new approach to variable metric methods: *Comput. J.*, **13**, 317–322.
- Goldfarb, D., 1970, A family of variable metric methods derived by variational means: *Math. Comp.*, **24**, 23–26.
- Guitton, A., and Symes, W. W., 1999, Robust and stable velocity analysis using the Huber function: 69th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1166–1169.
- Guitton, A., 2000, Huber solver versus IRLS algorithm for quasi L1 inversion: *SEP-103*, 255–271.
- Huber, P. J., 1973, Robust regression: Asymptotics, conjectures, and Monte Carlo: *Ann. Statist.*, **1**, 799–821.
- Huber, P. J., 1981, *Robust statistics*: Wiley series in Probability and Mathematical statistics.
- Kelley, C. T., 1999, *Iterative methods for optimization*: SIAM in applied mathematics.
- Li, W., 1995, Numerical algorithms for the Huber M-estimator problem: *Approximation Theory*, **8**, 1–10.
- Liu, D. C., and Nocedal, J., 1989, On the limited memory BFGS method for large scale optimization: *Mathematical Programming*, **45**, 503–528.
- More, J. J., and Thuente, J., 1994, Line search algorithms with guaranteed sufficient decrease: *ACM Transactions on Mathematical Software*, **20**, 286–307.
- Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: *Mathematics of Computation*, **95**, 339–353.
- Shanno, D. F., 1970, Conditioning of quasi-Newton methods for function minimization: *Math. Comp.*, **24**, 647–657.
- Tarantola, A., 1987, *Inverse Problem Theory: methods for data fitting and model parameter estimation*: Elsevier.

Helical preconditioning and splines in tension

Sergey Fomel¹

ABSTRACT

Splines in tension are smooth interpolation surfaces whose behavior in unconstrained regions is controlled by the tension parameter. I show that such surfaces can be efficiently constructed with recursive filter preconditioning and introduce a family of corresponding two-dimensional minimum-phase filters. The filters are created by spectral factorization on a helix.

INTRODUCTION

The method of minimum curvature is an old and ever-popular approach for constructing smooth surfaces from irregularly spaced data (Briggs, 1974). The surface of minimum curvature corresponds to the minimum of the Laplacian power or, in an alternative formulation, satisfies the biharmonic differential equation. Physically, it models the behavior of an elastic plane. In the one-dimensional case, the minimum curvature method leads to the natural cubic spline interpolation (de Boor, 1978). In the two-dimensional case, a surface can be interpolated with biharmonic splines (Sandwell, 1987) or gridded with an iterative finite-difference scheme (Swain, 1976). Claerbout (1999) suggests a straightforward least-squares optimization approach employing an iterative conjugate-gradient algorithm.

In most of the practical cases, the minimum curvature method produces a visually pleasing smooth surface. However, in cases of large changes in the surface gradient, the method can create strong artificial oscillations in the unconstrained regions. Switching to lower-order methods, such as minimizing the power of the gradient, solves the problem of extraneous inflections, but also removes the smoothness constraint and leads to gradient discontinuities (Fomel and Claerbout, 1995). A remedy, suggested by Schweikert (1966), is known as *splines in tension*. Splines in tension are constructed by minimizing a modified quadratic form that includes a tension term. Physically, the additional term corresponds to tension in elastic plates (Timoshenko and Woinowsky-Krieger, 1968). Smith and Wessel (1990) developed a practical algorithm of 2-D gridding with splines in tension and implemented it in the GMT software package.²

Fomel et al. (1997) have recently shown that an iterative interpolation algorithm can be greatly accelerated by preconditioning with recursive multidimensional filters defined on a

¹email: sergey@sep.stanford.edu

²<http://www.soest.hawaii.edu/gmt/>

helix (Claerbout, 1998a,b). To construct a minimum-phase filter suitable for recursive filtering, one can apply an efficient spectral factorization method (Sava et al., 1998).

In this paper, I develop an application of helical preconditioning to gridding with splines in tension. I introduce a family of 2-D minimum-phase filters for different degrees of tension. The filters are constructed by spectral factorization of the corresponding finite-difference forms. In the case of zero tension (the original minimum-curvature formulation), we obtain a minimum-phase version of the Laplacian filter. The case of infinite tension leads to spectral factorization of the Laplacian and produces the known *helical derivative* filter (Claerbout, 1999; Zhao, 1999).

The tension filters can be applied not only for interpolation but also for preconditioning in any estimation problems with smooth models. Tomographic velocity estimation is an obvious example of such an application (Woodward et al., 1998).

MATHEMATICAL THEORY OF SPLINES IN TENSION

The traditional minimum-curvature criterion implies seeking a two-dimensional surface $f(x, y)$ in region D , which corresponds to the minimum of the Laplacian power:

$$\iint_D \nabla^2 f(x, y) dx dy, \quad (1)$$

where ∇^2 denotes the Laplacian operator: $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$.

Alternatively, we can seek $f(x, y)$ as the solution of the biharmonic differential equation

$$(\nabla^2)^2 f(x, y) = 0. \quad (2)$$

Equation (2) corresponds to the normal system of equations in the least-square optimization problem. Briggs (1974) derives it directly from (1) with the help of Gauss's theorem.

Formula (1) approximates the strain energy of a thin elastic plate (Timoshenko and Woinowsky-Krieger, 1968). Taking tension into account modifies both the energy formula (1) and the corresponding equation (2). Smith and Wessel (1990) suggest the following form of the modified equation:

$$[(1-t)(\nabla^2)^2 - t(\nabla^2)] f(x, y) = 0, \quad (3)$$

where the tension parameter t ranges from 0 to 1. Zero tension leads to the biharmonic equation (2) and corresponds to the minimum curvature construction. The case of $t = 1$ corresponds to infinite tension. Although infinite tension is physically impossible, the resulting Laplace equation does have a physical interpretation of a steady-state temperature distribution. An important property of harmonic functions (solutions of the Laplace equation) is that they cannot have local minima and maxima in the free regions. With respect to interpolation, this means that, in the case of $t = 1$, the interpolation surface will be constrained to have its local extrema only at the input locations.

To interpolate an irregular set of data values, f_k at points (x_k, y_k) , we need to solve equation (3) under the constraint

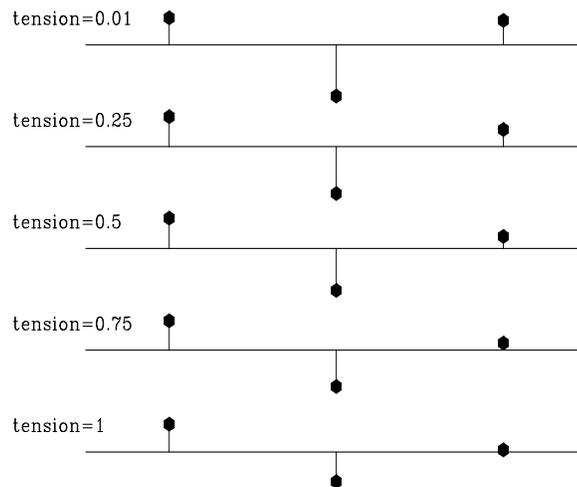
$$f(x_k, y_k) = f_k . \tag{4}$$

An iterative solution of this problem can be greatly accelerated by preconditioning (Fomel, 1997; Fomel et al., 1997). If \mathbf{A} is the discrete filter representation of the differential operator in equation (3), and we can find a minimum-phase filter \mathbf{D} whose autocorrelation is equal to \mathbf{A} , then an appropriate preconditioning operator is a recursive inverse filtering with the filter \mathbf{D} . Formulating the problem in helical coordinates (Claerbout, 1998a,b) allows us to perform both the spectral factorization of \mathbf{A} and inverse filtering with \mathbf{D} .

FINITE DIFFERENCES AND SPECTRAL FACTORIZATION

In the one-dimensional case, a finite-difference representation of the squared Laplacian can be defined as a centered 5-point filter with coefficients $(1, -4, 6, -4, 1)$. On the same grid, the Laplacian operator can be approximated to the same order of accuracy with the filter $(1/12, -4/3, 5/2, -4/3, 1/12)$. Combining the two filters in accordance with equation (3) and performing a spectral factorization with one of the standard methods (Claerbout, 1976, 1992), we can obtain a 3-point minimum-phase filter, suitable for inverse filtering. Figure 1 shows a family of one-dimensional minimum-phase filters for different values of the parameter t . Figure 2 demonstrates the interpolation results obtained with these filters on a simple one-dimensional synthetic. As expected, a small tension value ($t = 0.01$) produces a smooth interpolation, but creates artificial oscillations in the unconstrained regions around sharp changes in the gradient. The value of $t = 1$ leads to linear interpolation with no extraneous inflections, but with discontinuous derivatives. Intermediate values of t allow us to achieve a compromise: a smooth surface with constrained oscillations.

Figure 1: One-dimensional minimum-phase filters for different values of the tension parameter t . The filters range from the second derivative for $t = 0$ to the first derivative for $t = 1$. sergey1-otens [ER]



To design the corresponding filters in two dimensions, I define the finite-difference representation of operator (3) on a 5-by-5 stencil. The filters coefficients are chosen with the help

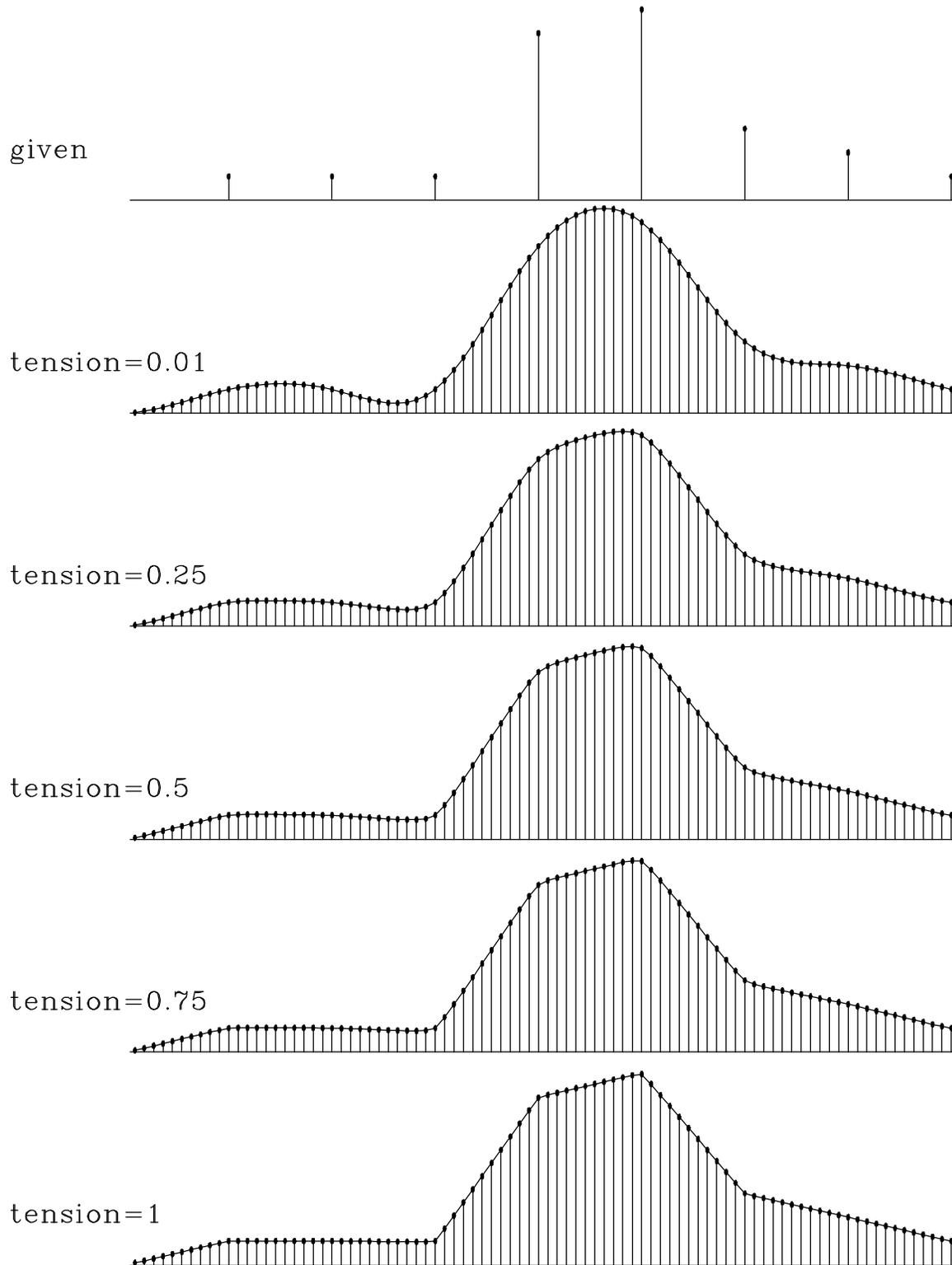


Figure 2: Interpolating a simple one-dimensional synthetic with recursive filter preconditioning for different values of the tension parameter t . The input data is shown on the top. The interpolation results range from a natural cubic spline interpolation for $t = 0$ to linear interpolation for $t = 1$. `sergey1-int [ER,M]`

of the Taylor expansion to match the desired spectrum of the operator around the zero spatial frequency. The matching conditions lead to the following set of coefficients for the squared Laplacian:

$$\begin{array}{ccccc}
 -1/60 & 2/5 & 7/30 & 2/5 & -1/60 \\
 2/5 & -14/15 & -44/15 & -14/15 & 2/5 \\
 7/30 & -44/15 & 57/5 & -44/15 & 7/30 \\
 2/5 & -14/15 & -44/15 & -14/15 & 2/5 \\
 -1/60 & 2/5 & 7/30 & 2/5 & -1/60
 \end{array} = 1/60 \begin{array}{ccccc}
 -1 & 24 & 14 & 24 & -1 \\
 24 & -56 & -176 & -56 & 24 \\
 14 & -176 & 684 & -176 & 14 \\
 24 & -56 & -176 & -56 & 24 \\
 -1 & 24 & 14 & 24 & -1
 \end{array}$$

Laplacian representation with the same order of accuracy has the coefficients

$$\begin{array}{ccccc}
 -1/360 & 2/45 & 0 & 2/45 & -1/360 \\
 2/45 & -14/45 & -4/5 & -14/45 & 2/45 \\
 0 & -4/5 & 41/10 & -4/5 & 0 \\
 2/45 & -14/45 & -4/5 & -14/45 & 2/45 \\
 -1/360 & 2/45 & 0 & 2/45 & -1/360
 \end{array} = 1/360 \begin{array}{ccccc}
 -1 & 16 & 0 & 16 & -1 \\
 16 & -112 & -288 & -112 & 16 \\
 0 & -288 & 1476 & -288 & 0 \\
 16 & -112 & -288 & -112 & 16 \\
 -1 & 16 & 0 & 16 & -1
 \end{array}$$

For the sake of simplicity, I assumed an equal physical spacing in x and y directions. The coefficients can be easily adjusted for anisotropic spacing. Figures 3 and 4 show the spectra of the finite-difference representations of operator (3) for the different values of the tension parameter. The finite-different spectra appear as fairly isotropic. They match the exact expressions at small frequencies.

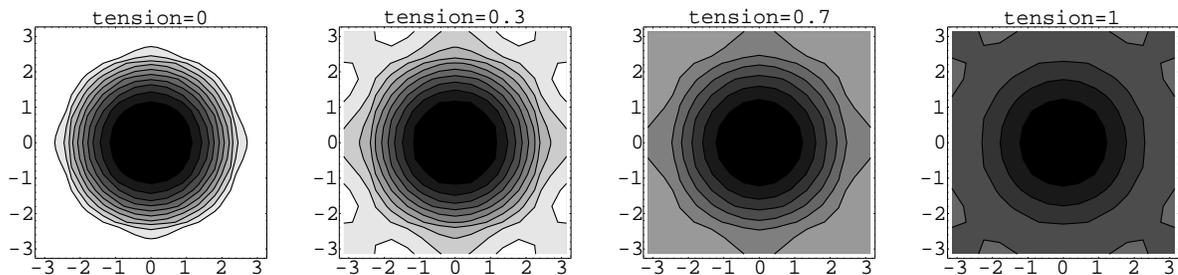


Figure 3: Spectra of the finite-difference splines-in-tension schemes for different values of the tension parameter (contour plots). `sergey1-specc` [CR]

Regarding the finite-difference operators as two-dimensional auto-correlations and applying the efficient Wilson-Burg method of spectral factorization (Claerbout, 1999; Sava et al., 1998), I obtain two-dimensional minimum-phase filters suitable for inverse filtering. The exact filters contain many coefficients, which rapidly decrease in magnitude at a distance from the first coefficient. For reasons of efficiency, it is advisable to restrict the shape of the filter so that it contains only the valuable coefficients. Keeping all the coefficients that are 1000 times smaller in magnitude than the leading coefficient creates a 53-point filter for $t = 0$ and a 35-point filter for $t = 1$, with intermediate filter lengths for intermediate values of t . When

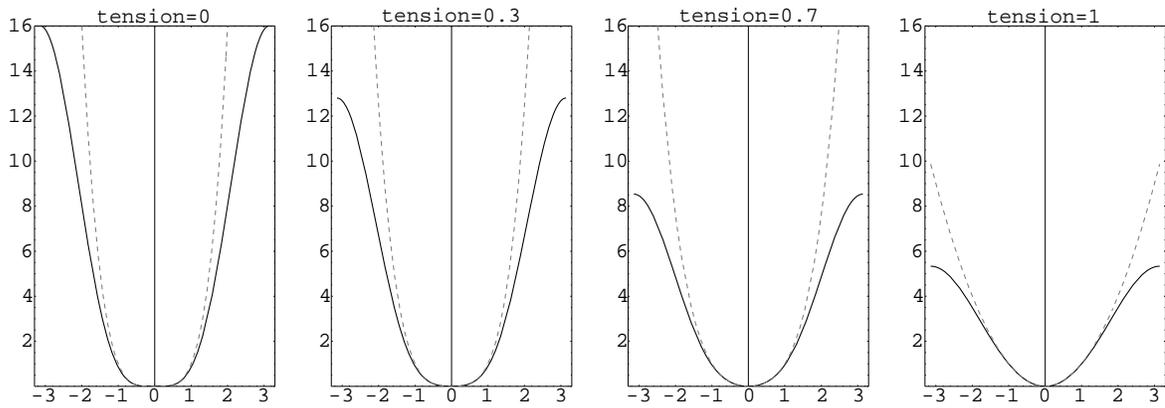


Figure 4: Spectra of the finite-difference splines-in-tension schemes for different values of the tension parameter (cross-section plots). Dashed lines show the exact spectra for continuous operators. [sergey1-specp](#) [CR]

the ratio is changed to 200, we obtain 25- and 16-point filters, respectively. The restricted filters don't factor the autocorrelation exactly, but provide an effective approximation of the exact factors. As outputs of the Wilson-Burg spectral factorization process, they obey the minimum-phase condition.

Figure 5 shows the two-dimensional filters for different values of t and illustrates inverse recursive filtering, which is the essence of the helix method (Claerbout, 1999, 1998a,b). The case of $t = 1$ leads to the filter known as *helix derivative* (Claerbout, 1999; Zhao, 1999). The filter values are spread mostly on two columns. The other boundary case of $t = 0$ leads to a three-column filter, which serves as the minimum-phase version of the Laplacian. As expected from the theory, the inverse impulse response of this filter is noticeably smoother and wider than the inverse response of the helix derivative. Filters corresponding to intermediate values of t exhibit intermediate properties. Theoretically, the inverse impulse response of the filter corresponds to the Green function of equation (3). The theoretical Green function for the case of $t = 1$ is

$$G = \frac{1}{2\pi} \ln r, \quad (5)$$

where r is the distance from the impulse: $r = \sqrt{(x - x_k)^2 + (y - y_k)^2}$. In the case of $t = 0$, the Green function is smoother at the origin:

$$G = \frac{1}{8\pi} r^2 \ln r. \quad (6)$$

The theoretical Green function expression for an arbitrary value of t is not known, but we can assume that its smoothness lies between the two boundary conditions.

In the next section, I illustrate an application of helical inverse filtering to a two-dimensional interpolation problem.

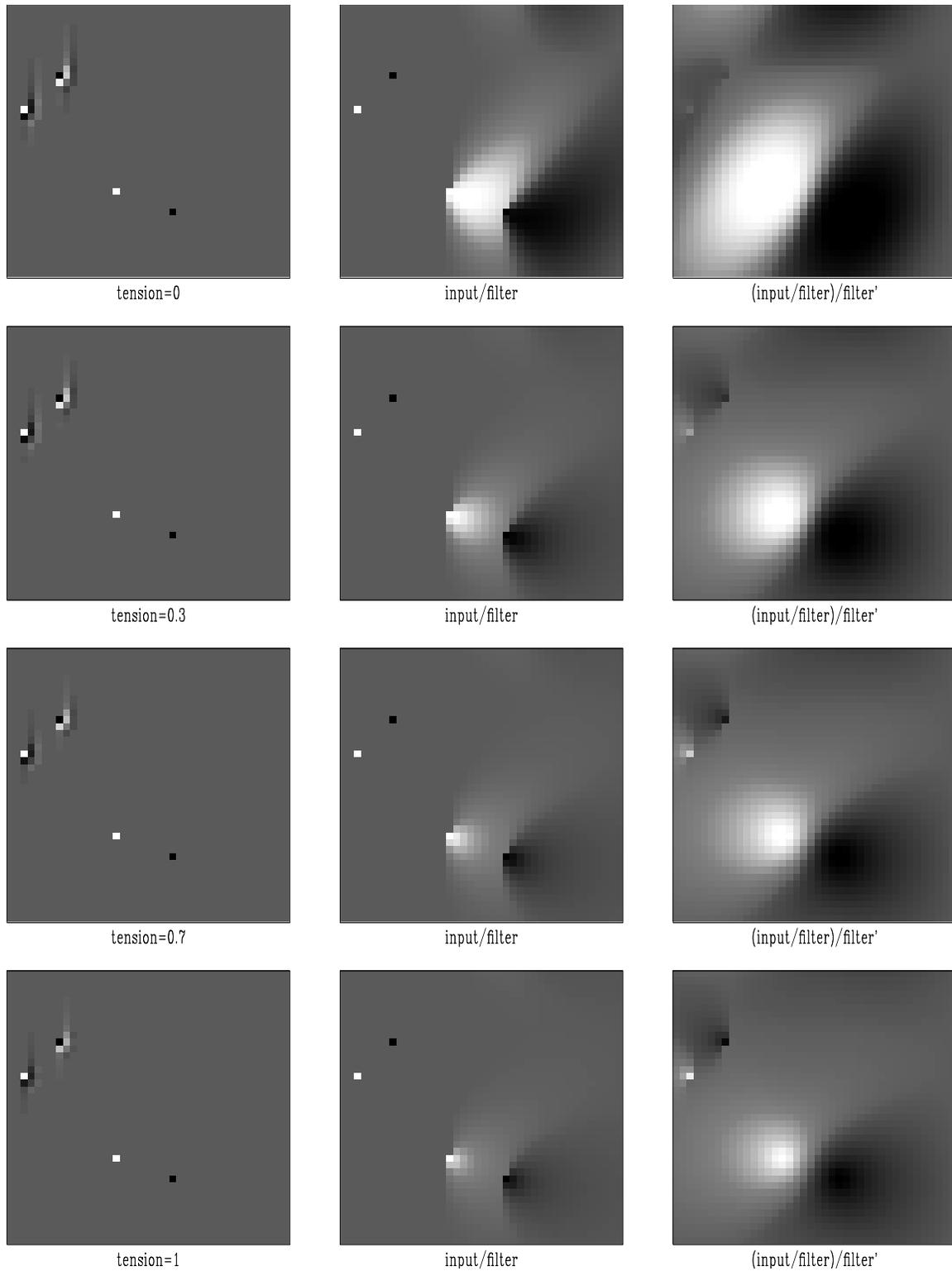


Figure 5: Inverse filtering with the tension filters. The left plots show the inputs composed of filters and spikes. Inverse filtering turns filters into impulses and turns spikes into inverse filter responses (middle plots). Adjoint filtering creates smooth isotropic shapes (right plots). The tension parameter takes values 0, 0.3, 0.7, and 1 (from top to bottom). `sergey1-splin` [ER,M]

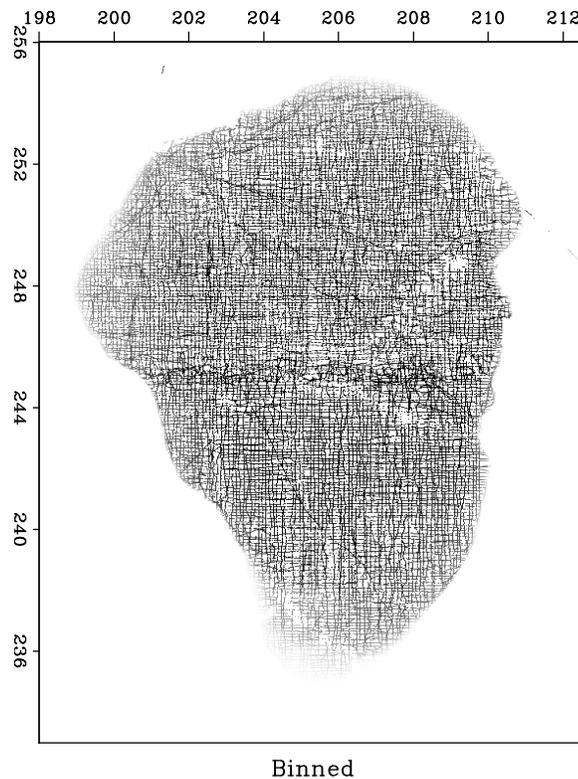
INTERPOLATION EXAMPLE

I chose the familiar Galilee dataset (Fomel and Claerbout, 1995; Claerbout, 1999) for a simple interpolation illustration. The data was collected on a bottom sounding survey of the Sea of Galilee in Israel (Ben-Avraham et al., 1990). The data contain a number of noisy, erroneous and inconsistent measurements, which present a challenge for the traditional estimation methods. Addressing this challenge completely goes beyond the scope of this paper.

Figure 6 shows the data after a nearest-neighbor binning to a regular grid. The data was then passed to an interpolation program to fill the empty bins. The results (for different values of t) are shown in Figures 7 and 8. Interpolation with the minimum-phase Laplacian ($t = 0$) creates a relatively smooth interpolation surface but plants artificial little mountains around the edge of the sea. This effect is caused by large gradient changes and is similar to the sidelobe effect in the one-dimensional example (Figure 2). It is clearly seen in the cross-section plots in Figure 8. Interpolation with the helix derivative ($t = 1$) is free from the sidelobe artifacts, but it also produces an undesirable non-smooth behavior in the middle part of the image. As in the one-dimensional example, intermediate tension allows us to achieve a compromise: smooth interpolation in the middle and constrained behavior at the sides of the sea bottom.

Figure 6: The Sea of Galilee dataset after a nearest-neighbor binning. The binned data is used as an input for the missing data interpolation program.

`sergey1-mesh` [ER]



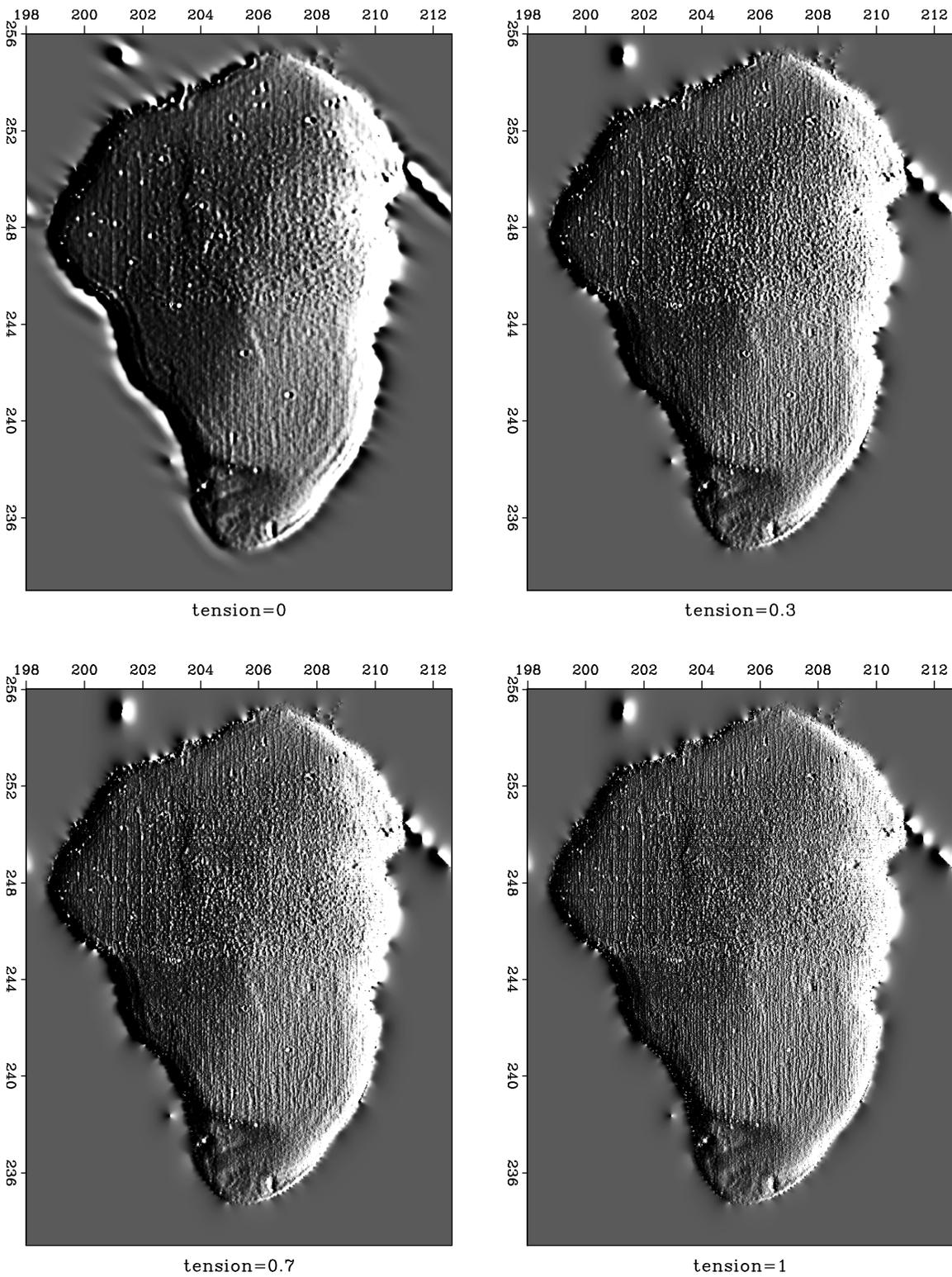


Figure 7: The Sea of Galilee dataset after missing data interpolation with helical preconditioning. Different plots correspond to different values of the tension parameter. An east-west derivative filter was applied to illuminate the surface. `sergey1-gal` [ER,M]

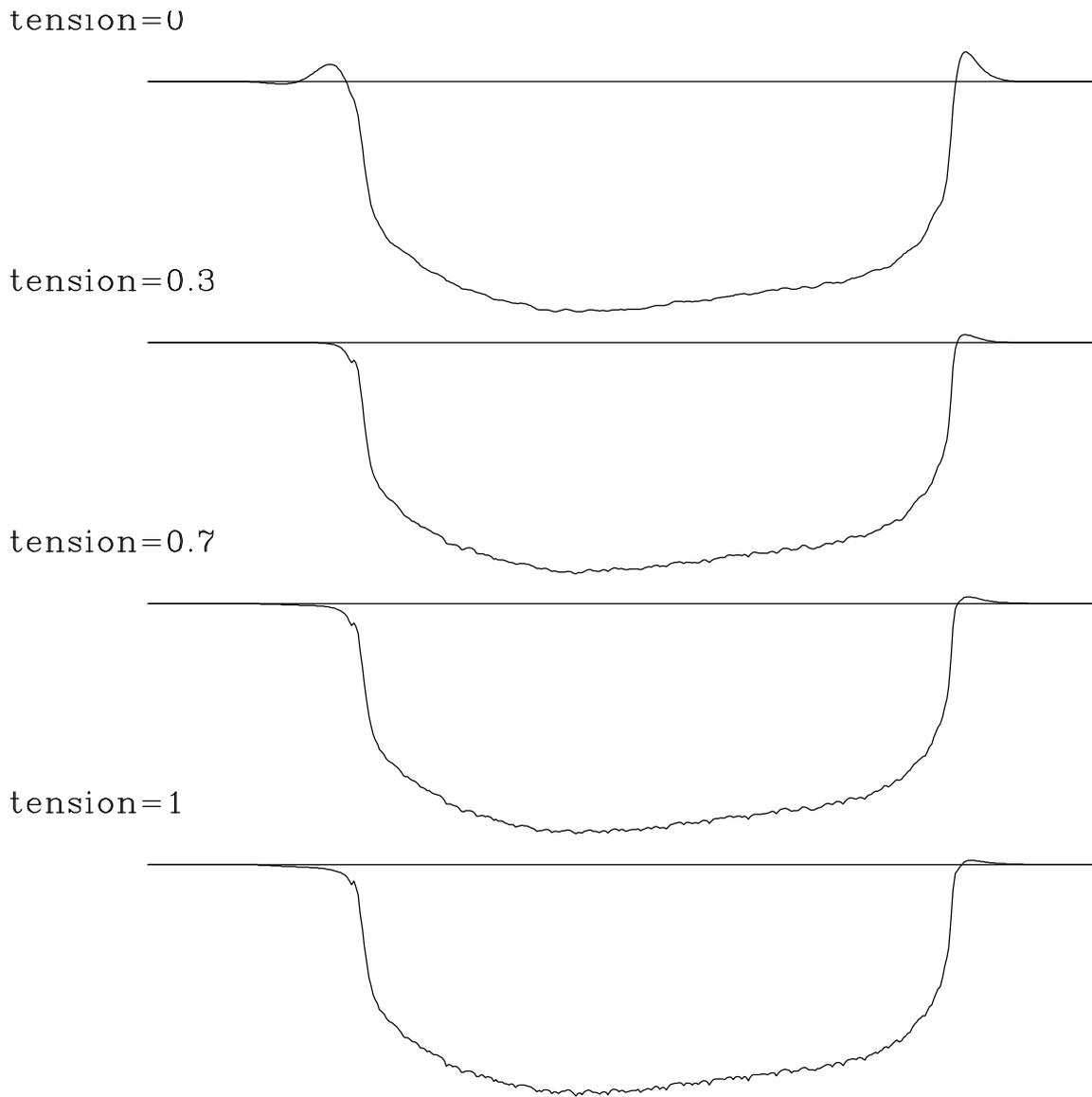


Figure 8: Cross-sections of the Sea of Galilee dataset after missing data interpolation with helical preconditioning. Different plots correspond to different values of the tension parameter.

sergey1-cross [ER]

CONCLUSIONS

Splines in tension represent an approach to constrained interpolation of smooth surfaces. The constraint is embedded in a user-specified tension parameter. The two boundary values of tension correspond to cubic and linear interpolation.

By applying the method of spectral factorization on a helix, I have been able to define a family of two-dimensional minimum-phase filters, which correspond to the spline interpolation problem with different values of tension. These filters contribute to our collection of useful helical filters. They can be used for preconditioning interpolation problems with smooth surfaces and, in general, for preconditioning geophysical estimation problems with smooth models.

REFERENCES

- Ben-Avraham, Z., Amit, G., Golan, A., and Begin, Z. B., 1990, The bathymetry of Lake Kinneret and its structural significance: *Israel Journal of Earth Sciences*, **39**, 77–84.
- Briggs, I. C., 1974, Machine contouring using minimum curvature: *Geophysics*, **39**, no. 1, 39–48.
- Claerbout, J. F., 1976, *Fundamentals of geophysical data processing*: Blackwell.
- Claerbout, J. F., 1992, *Earth Soundings Analysis: Processing Versus Inversion*: Blackwell Scientific Publications.
- Claerbout, J., 1998a, Multidimensional recursive filters via a helix: *SEP-97*, 319–335.
- Claerbout, J., 1998b, Multidimensional recursive filters via a helix: *Geophysics*, **63**, 1532–1541.
- Claerbout, J., 1999, *Geophysical estimation by example: Environmental soundings image enhancement: Stanford Exploration Project*, <http://sepwww.stanford.edu/sep/prof/>.
- de Boor, C., 1978, *A practical guide to splines*: Springer-Verlag.
- Fomel, S., and Claerbout, J., 1995, Searching the Sea of Galilee: The splendors and miseries of iteratively reweighted least squares: *SEP-84*, 259–270.
- Fomel, S., Clapp, R., and Claerbout, J., 1997, Missing data interpolation by recursive filter preconditioning: *SEP-95*, 15–25.
- Fomel, S., 1997, On model-space and data-space regularization: A tutorial: *SEP-94*, 141–164.
- Sandwell, D. T., 1987, Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter data: *Geophys. Res. Letters*, **14**, no. 2, 139–142.

- Sava, P., Rickett, J., Fomel, S., and Claerbout, J., 1998, Wilson-Burg spectral factorization with application to helix filtering: *SEP-97*, 343–351.
- Schweikert, D. G., 1966, An interpolation curve using a spline in tension: *Journal of Mathematics and Physics*, **45**, 312–313.
- Smith, W. H. F., and Wessel, P., 1990, Gridding with continuous curvature splines in tension: *Geophysics*, **55**, no. 3, 293–305.
- Swain, C. J., 1976, A FORTRAN IV program for interpolating irregularly spaced data using the difference equations for minimum curvature: *Computers and Geosciences*, **1**, 231–240.
- Timoshenko, S., and Woinowsky-Krieger, S., 1968, *Theory of plates and shells*: McGraw-Hill.
- Woodward, M. J., Farmer, P., Nichols, D., and Charles, S., 1998, Automated 3-D tomographic velocity analysis of residual moveout in prestack depth migrated common image point gathers: 68th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1218–1221.
- Zhao, Y., 1999, Helix derivative and low-cut filters' spectral feature and application: *SEP-100*, 235–250.

Patching and micropatching in seismic data interpolation

Sean Crawley¹

ABSTRACT

I interpolate CMP gathers with PEFs arranged on a dense, radial grid. The radial grid facilitates preconditioning by radial smoothing, and enables the use of relatively large grid cells, which we refer to as micropatches. Even when the micropatches contain enough data samples that the PEF calculation problem appears overdetermined, radial smoothing still noticeably improves the interpolation, particularly on noisy data.

INTRODUCTION

In filtering applications, input seismic data are commonly divided into smaller subsets which we refer to as patches (which are also referred to as windows, gates, and other things). The data are assumed to be approximately stationary within a patch, but due to practical limits on patch size, it may not be possible to avoid nonstationarity and poor results in some patches. This is often true where the data are strongly curved (for spatial filtering) or noisy. An alternative to independent patches is nonstationary filtering. SEP has applied nonstationary filtering to numerous problems in recent reports, including groundroll suppression (Brown et al., 1999), multiple suppression (Clapp and Brown, 1999), tomography regularization (Clapp and Biondi, 1998), deconvolution (Claerbout, 1997), and interpolation (Fomel, 1999; Crawley, 1999). Nonstationary PEFs do not have patch-size limitations, so we may shrink patches down to arbitrary size and shape. To control the potentially huge null space, we regularize the set of filters to ensure that PEFs located at similar data coordinates have similar coefficients. This implements the assumption that dips in the data may change everywhere, but do so smoothly. Besides being small, these new patches are fundamentally different in that they are not independent problems, but related to each other via the regularization. To distinguish them from the old independent patches, we call them “micropatches”. We have a great deal of freedom in deciding the size and shape of our micropatches, and in implementing the regularization. This paper motivates and describes my implementation.

¹email: sean@sep.stanford.edu

FORMULATION

A standard formulation for calculating PEFs from known data is to solve a linear least-squares problem like

$$\mathbf{0} \approx \mathbf{YCa} + \mathbf{r}_0, \quad (1)$$

where \mathbf{a} is a vector containing the PEF coefficients, \mathbf{C} is a filter coefficient selector matrix, and \mathbf{Y} denotes convolution with the input data. The coefficient selector \mathbf{C} is like an identity matrix, with a zero on the diagonal placed to prevent the fixed 1 in the zero lag of the PEF from changing. The \mathbf{r}_0 is a vector that holds the initial value of the residual, \mathbf{Ya}_0 . If the unknown filter coefficients are given initial values of zero, then \mathbf{r}_0 contains a copy of the input data. \mathbf{r}_0 makes up for the fact that the 1 in the zero lag of the filter is not included in the convolution (it is knocked out by \mathbf{C}). When there are many coefficients, as when PEFs are spread densely on the data grid, it makes sense to add damping equations and/or precondition the problem. Inserting the preconditioned variable \mathbf{Sp} (where \mathbf{S} is a somewhat arbitrary smoother) for \mathbf{a} and adding the also somewhat arbitrary roughener $\mathbf{R} = \mathbf{S}^{-1}$ to regularize the model, gives a formulation like

$$\mathbf{0} \approx \mathbf{YKSp} + \mathbf{r}_0 \quad (2)$$

$$\mathbf{0} \approx \epsilon \mathbf{Ip} \quad (3)$$

In many cases we can set $\epsilon = 0$ and just use equation (2), being careful not to let it go for too many iterations. We still have to define \mathbf{S} (or \mathbf{R}).

RADIAL SMOOTHING

We have to choose \mathbf{S} . Inserting a smoother signifies the assertion that the dips in seismic data should change in a gradual way. Choosing an isotropic smoother means we expect the dips to vary similarly in all directions. However, we know that the dip spectrum of the data probably changes more quickly in some directions than in others. We want to smooth most heavily along directions where the dip is nearly constant. In a constant-velocity, flat-layered earth, events fall along hyperbolas like

$$t^2 = \tau^2 + \frac{x^2}{v^2},$$

where x is offset, v is stacking velocity, τ is zero-offset time. The time dip of an event is dt/dx . If velocity is constant, differentiating gives

$$\frac{dt}{dx} = \frac{x}{v^2 t},$$

which means that the dip does not change along radial lines, where x/t is constant. In a real earth, we suppose that dips will change, but slowly. Real earth velocity may change quickly in depth, but hyperbola trajectories are functions of RMS velocity, which is smooth. We want a smoother with an impulse response which is highly elongated in the radial direction. To get

a big impulse response cheaply, I apply the inverse of a directional derivative, pointed in the radial direction. To directly apply the inverse, the roughener has to be causal, which means that the inverse will only smooth in one direction (Claerbout, 1998). We want \mathbf{S} to have an impulse response which is smoothed both in towards zero radius and out towards large radius, so we make it the cascade of the causal smoother and its anticausal adjoint.

Patches, micropatches, pixels

Having chosen the radial direction, we can think of some different ways of implementing our radially-smoothed filters. An obvious one is putting a PEF at every point on the data grid, and devising a derivative filter which adjusts its direction to point at the origin. An alternative is to overlay a radial grid on the data grid, and arrange PEFs on the radial grid. Here we compare the two smoothing schemes. Our goal is to assume stationarity in a small enough region that we can interpolate well where the data do not fit a plane-wave model. In the method of independent patches, individual patches are treated as separate problems. A patch can not be arbitrarily small, because it must provide enough fitting equations that the filter coefficients are well overdetermined. In 1-D, filtering with a PEF looks like this:

$$\hat{u}(i) = \sum_{k=0}^p a(k)u(i-k), \quad p + i_{\text{patchmin}} \leq i \leq i_{\text{patchmax}}. \quad (4)$$

The patch boundaries are i_{patchmin} and i_{patchmax} , \mathbf{p} is the number of adjustable coefficients, $a(0) = 1$. The lower limit on i is to prevent the filter from running off the end of the data and encountering implicit zero values. The patches are designed to overlap, and the outputs are normalized to hide the patch boundaries. In moving to the method of gradually-varying PEFs, we replace the notion of extracting a subset of the data with that of dereferencing the data coordinates to find the appropriate filter, as in

$$\hat{u}(i) = \sum_{k=0}^p a_i(k)u(i-k), \quad p + 1 \leq i \leq i_{\text{datamax}}. \quad (5)$$

The index of the data sample i dereferences the set of filters $a_i(k)$. The data boundaries $i = 1$ and i_{datamax} replace the patch boundaries. We have lots of freedom in dereferencing a_i . In the limiting cases, all the data may share one PEF, or we can choose a_i to be a different set of coefficients for each data point. In the case where we have a PEF at every data point, we call \mathbf{S} pixel-wise smoothing. Choosing a separate PEF for every input sample is a possibility, but not necessary. Our motivation for moving away from independent patches was to use one PEF in a region small enough that we do not have trouble with nonstationarity. Some amount of patching may still make sense, provided the patches may be small. It is easy to implement small patches as a generalization of the case above where each data sample has its own PEF. A particular a_i can be the same for any number of values of i without complication. Because they may be small, we refer to the new patches as “micropatches” to distinguish between them and independent patches. To subdivide a CMP gather into micropatches, we choose a web-like grid made up of radial lines and circular lines. Radial lines are a natural choice because we

want to smooth in the radial direction. Circles are somewhat arbitrary; we could choose flat lines or reflection-like hyperbolas to cross the radial lines. Circles have the attractive property that they make equal-area micropatches at a given radius.

Smoothing pixels versus smoothing micropatches

We can choose pixel-wise smoothing or micropatch smoothing. An easy argument favoring micropatches over pixel-wise smoothing says that putting a filter at every data sample is a tremendous waste of memory. If the data are predictable at all, they are probably not so nonstationary that they need a separate PEF at each sample. A single 3-D PEF has easily 20 or more adjustable coefficients, so allocating the set of PEFs requires 20 times the storage of the input data. Even very small micropatches require much less memory. Micropatches also have some simplifying side effects that make them preferable to pixel-wise smoothing. One is apparent from examining Figures 1 and 2. Figure 1 shows smoothing in micropatches and Figure 2 shows pixel-wise smoothing. In each figure, the values represent filter coefficients displayed in data coordinates. The axes are time and offset. The top halves show a set of impulses, labeled \mathbf{d} . \mathbf{Md} is the impulses binned into micropatches, while \mathbf{Pd} is the impulses binned into pixels (naturally, $\mathbf{Pd} = \mathbf{d}$). \mathbf{F} and \mathbf{F}' are pixel-wise smoothers pointed towards and away from zero radius, respectively. \mathbf{C} and \mathbf{C}' are the micropatched smoothers. In this case, the two are similar, though the pixel-wise smoother obviously produces a higher-resolution picture (though the micropatches could be made much smaller). The bottom halves show the same treatment applied to a constant function, labeled $\mathbf{1}$. $\mathbf{M1}$ has an angular limit applied. Pixel-wise smoothing creates some very large ridge artifacts, visible in $\mathbf{FP1}$ and $\mathbf{F'FP1}$, where the angle between a data sample and the origin corresponds to an integer slope. Also, where the constant function is smoothed in towards zero radius, $\mathbf{FP1}$, energy concentrates in a huge spike at the origin. $\mathbf{F'F}$ and $\mathbf{C'C}$ can be thought of as weighting functions in equation (2) (either $\mathbf{F'F}$ or $\mathbf{C'C}$ is used for \mathbf{S} in that equation). It is desirable to have the flatter weighting function. It is also simpler to implement. Producing the many different angles in Figure 2 requires that the smoothers \mathbf{F} and \mathbf{F}' be made up of many different filters, oriented in a continuous sweep between a spatial derivative and a time derivative. \mathbf{C} and \mathbf{C}' produce the same range of angles in Figure 1 using a single, radial derivative filter. The PEFs in micropatches are regularly gridded in angle and radius, so they are easily smoothed in those directions with old-fashioned stationary 1-D derivative filters. PEFs at every pixel are instead regularly sampled in time and offset, so working in polar coordinates requires some work. \mathbf{F} uses many coefficients, \mathbf{C} uses two.

Is smoothing necessary?

Using the weblike pattern seen in Figure 1, it is often possible to use fairly large micropatches. An alternative to radial smoothing may be to simply lengthen the micropatches in the radial direction, and not bother smoothing at all. I test this in Figures 3 and 4. It seems that smoothing may have some important effects beyond just statistically compensating for the small size of a micropatch. Even with very elongated patches, such that the area of a patch is more than large

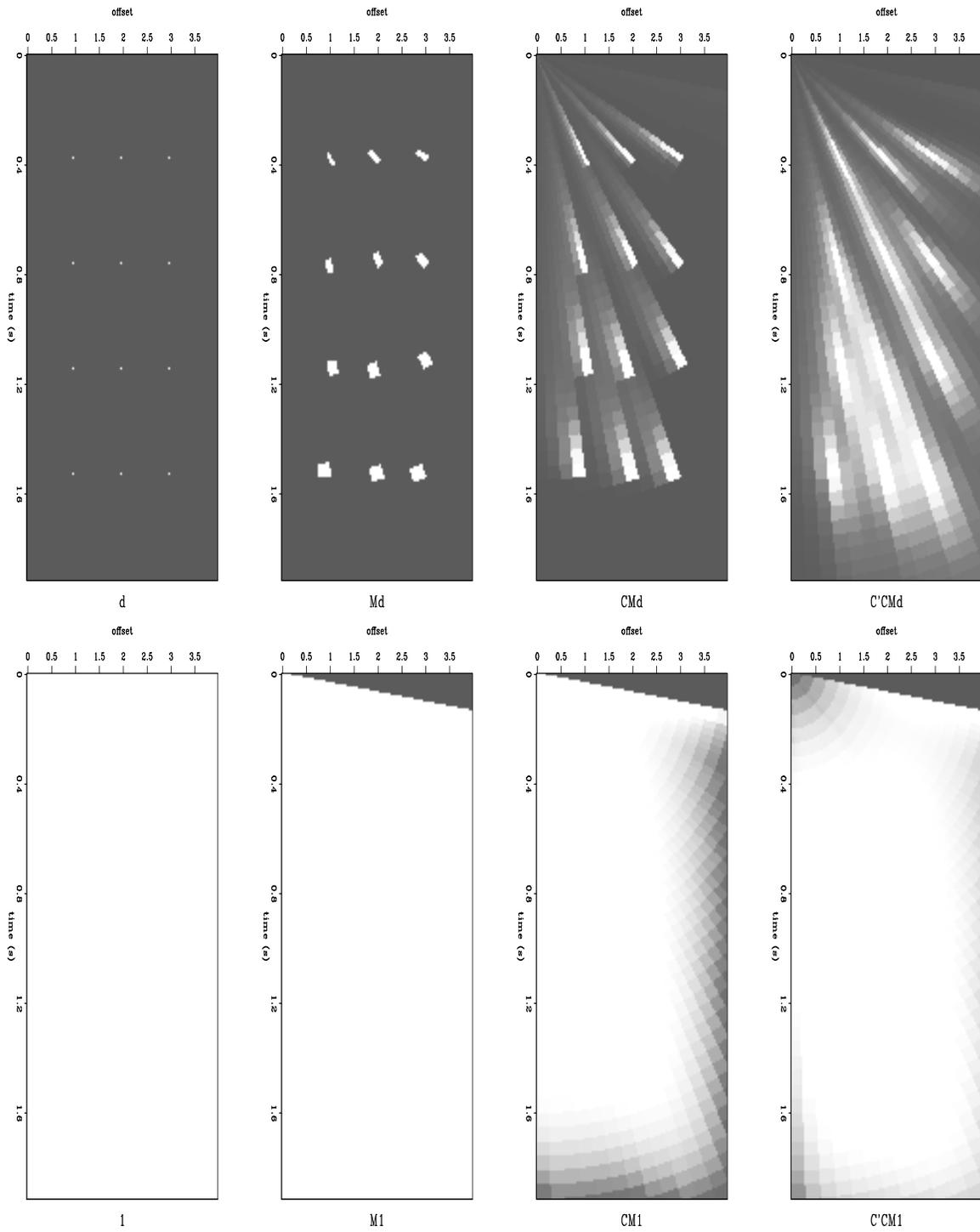


Figure 1: Illustration of micropatched radial filter coefficient smoothing. sean1-curtSmear8
[ER]

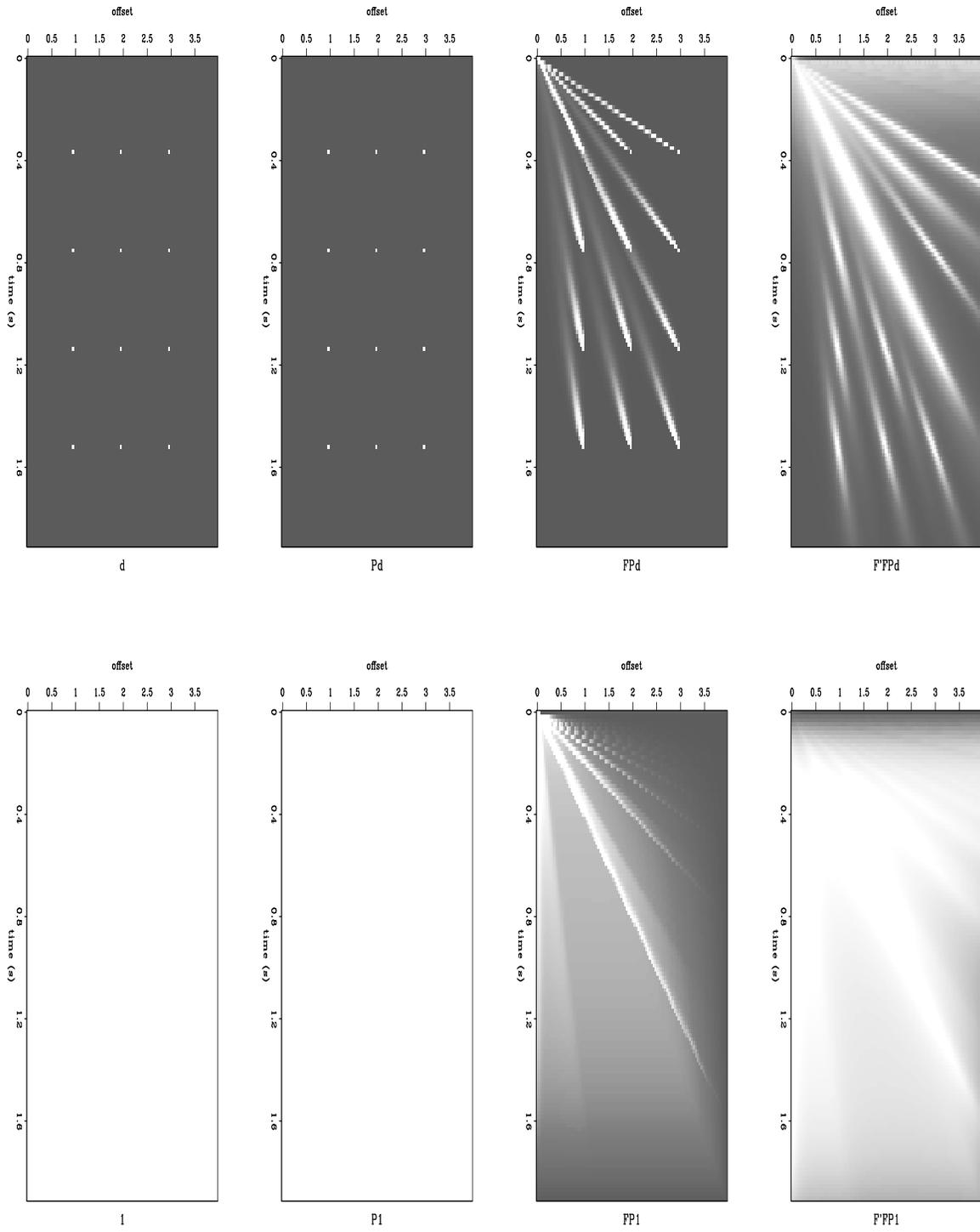


Figure 2: Illustration of pixel-wise radial filter coefficient smoothing. sean1-random8 [ER]

enough for the number of adjustable filter coefficients, smoothing noticeably improves the final result, particularly where the data have many dips or are noisy. One possible explanation is that where the data are incoherent, the change in a particular filter coefficient at each iteration is just an average of data samples, which is approximately zero. With the addition of the smoother, the change in nearby filter coefficients fills in. Figure 3 shows a section of noise-free data with many dips, interpolated with PEF smoothing on the left and without on the right. The top two panels show the interpolated traces (the known input traces are windowed out). The bottom two panels show the differences between the interpolated traces and the original traces which were thrown out to make the input. The two panels are similar, though the left side is noticeably better on some events. Figure 4 shows two more interpolation results. In this case the data is land data, and much noisier. Both known and interpolated traces are shown. Because the data is somewhat noisy, it is easier to distinguish between the coherency of the two panels than picking out differences between particular events. The result using PEF smoothing, in the left panel, is noticeably more coherent, particularly between 1.2 and 1.6 seconds.

CONCLUSION

We have a great deal of freedom in choosing how to distribute PEFs in the data coordinates, and in choosing how to implement the smoother between PEFs. Choosing the web-like arrangement of micropatches and smoothing PEFs in radial coordinates gives a nice flat smoother response, and the ability to use relatively large micropatches without trouble related to nonstationarity. The large size of the micropatches calls into question whether smoothing is actually necessary. It turns out that, especially where data are noisy, smoothing continues to be important.

REFERENCES

- Brown, M., Clapp, R. G., and Marfurt, K., 1999, Predictive signal/noise separation of groundroll-contaminated data: *SEP-102*, 111–128.
- Claerbout, J. F., 1997, *Geophysical exploration by example: Environmental soundings image enhancement*: Stanford Exploration Project.
- Claerbout, J. F., 1998, Multi-dimensional recursive filtering via the helix: *Geophysics*, **63**, no. 5, 1532–1541.
- Clapp, R. G., and Biondi, B. L., 1998, Regularizing time tomography with steering filters: *SEP-97*, 137–146.
- Clapp, R. G., and Brown, M., 1999, Applying sep's latest tricks to the multiple suppression problem: *SEP-102*, 91–100.
- Crawley, S., 1999, Interpolation with smoothly nonstationary prediction-error filters: *SEP-100*, 181–196.

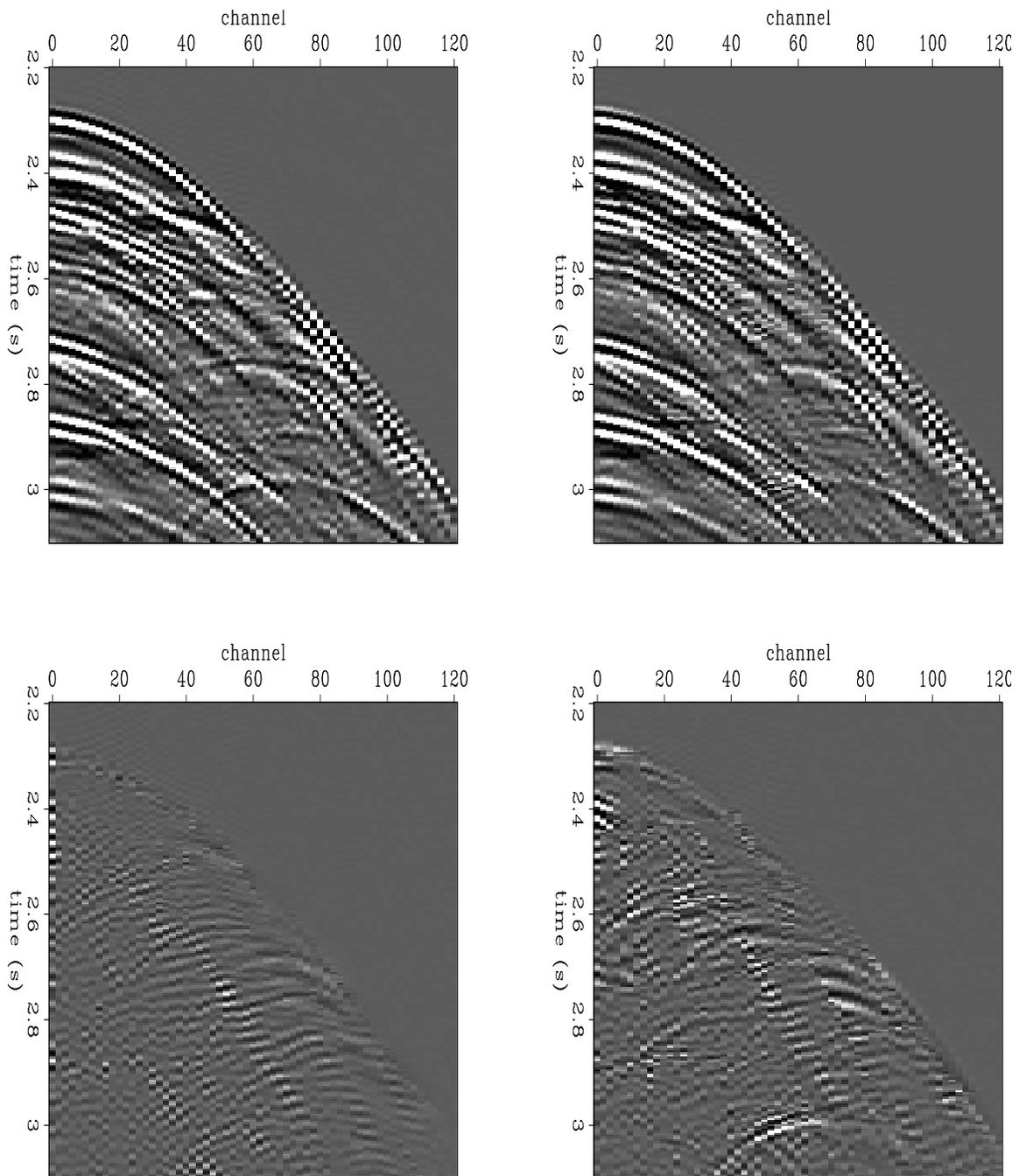


Figure 3: Noise-free interpolated traces and difference from original data. Traces were interpolated with PEF smoothing on the left, without on the right. Known data is not shown, to make the differences easier to see. The results are similar, but the result with smoothing is better. `sean1-smonosmo` [ER]

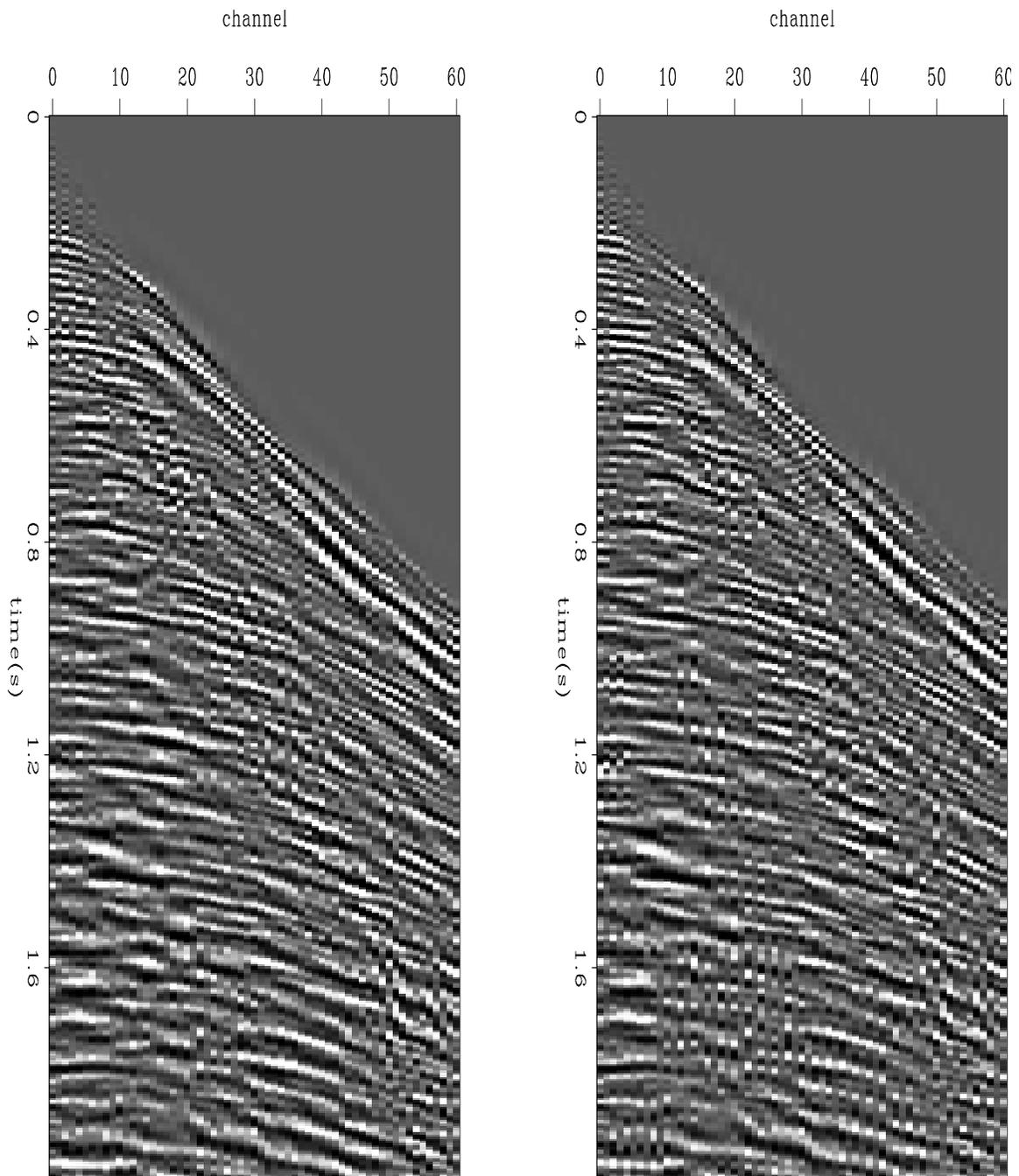


Figure 4: Noisy interpolated traces and differences. Traces were interpolated with PEF smoothing on the left, without on the right. Differences are more visible here on noisy data than on noise-free data. Traces interpolated without PEF smoothing abruptly change from coherent to incoherent along certain micropatch boundaries. [sean1-smonosmo2](#) [ER]

Fomel, S., 1999, Plane wave prediction in 3-D: SEP-**102**, 101-110.

Short Note

The Burg Method on a Helix?

Jon Claerbout¹

We have not yet put the Burg method of PEF estimation on a helix. The first reason to do so is that the Burg method assures us a stable PEF. The second reason to do so is that the Burg method should be much faster than conjugate gradients. For data length ND and filter length NF, the PEF estimation costs are

Conjugate gradients	ND*NF**2
Levinson	ND*NF + NF**2
Burg	ND*NF

Estimating PEF's on a helix with the Burg method does not seem difficult: Terms in sums that involve missing data can simply be omitted from certain averages. We could probably proceed much as we now do with conjugate gradients (CG).

PEF estimation is not our main problem, however. Our main problem is missing data. The Burg method has not yet been adapted to missing data estimation but we should try.

It remains to be seen how we can estimate missing values, both off the ends of the data and internal to it. As with CG, polynomial division seems to be an important part of the solution.

BURG PEF ESTIMATION REVIEW

Burg PEF estimation should work fine on a helix. Full details along with the 1-D code are found at (Claerbout, 1976). I will quickly review the theory from memory (partly to see how simple I can make it).

First is the notion that PEFs can be built up from this recursion

$$A^{N+1}(Z) = A^N(Z) + cZ^{N+1}A^N(1/Z) \quad (1)$$

where c is in the range $-1 \leq c \leq +1$. There is a theorem from Algebra that is easy to prove that if c is in the required range, then $A^{N+1}(Z)$ will be minimum phase if $A^N(Z)$ is minimum phase. Since $A^0 = 1$, all are minimum phase.

Burg's PEF calculation begins from two copies of the data $X(Z)$. One, called $F(Z)$ will be turned into the forward prediction error $A(Z)X(Z)$. The other called $B(Z)$ will be turned

¹email: claerbout@stanford.edu

into the backward prediction error $A(1/Z)X(Z)$. At each stage of the calculation, we compute c with this formula

$$c = -2 \frac{b \cdot f}{b \cdot b + f \cdot f} \quad (2)$$

The triangle inequality shows that for arbitrary f and b , c is in the required range. Given c we now form an upgraded F and B with $f \leftarrow f + cb$ and $b \leftarrow b + cf$. [At successive iterations, increasing time lag is introduced between f and b . Details in (1976).] When you are finished, you have $F = A(Z)X(Z)$ and $B = A(1/Z)X(Z)$. Why is that? To understand that requires delving into 1-D theory, in particular, the Levinson recursion, and we won't do that now. [Maybe I can think up an easier explanation later. Perhaps by a sequence of orthogonality arguments.] I recall if you append a tiny impulse function off the end of X before you start, when you finish, you will see it has turned into the PEF.

Now let us think about missing data off the ends of the Burgian one-dimensional data set. Given that we have computed $F(Z) = A(Z)X(Z)$, then we should find that $F(Z)/A(Z)$ matches $X(Z)$ until its end, and it is a logical continuation thereafter. Likewise $B(Z)$ could be used for extensions before the beginning of $X(Z)$. Thus it remains to think about how to handle gaps in the middle.

BURG PEF ESTIMATION ON A HELIX

Now, how does the Burg method fit on a helix? There is nothing new except for the huge gap while we wind around the back of the helix. In this gap, we would simply presume $c = 0$ and we do nothing there. We compute the PEF and the prediction error simply by omitting steps that we would ordinarily do.

If, however, we intend to use the PE filter, then we have some details to attend to, and this begins to get complicated. Reviewing the Levinson recursion, we find that gaps internal to the filter tend to fill as the recursion proceeds. The filter is not as sparse as the reflection coefficients. We'll need to keep track of the nonzero filter coefficients. We need to keep track of them in order that we have a PEF that we can use in polynomial division because polynomial division is an essential part of finding missing data with the Burg method and polynomial division is a part of preconditioning the conjugate-gradient method.

A promising thought is that perhaps the Burg recursion can be run backwards. Since this would take a PEF (or its reflection coefficients) and white inputs (forward and backward prediction errors) and create a colored outputs, it seems analogous to polynomial division.

When I began multidimensional filtering studies I was ignorant of the helix and thus had not the opportunity to use the Burg or the Levinson methods. Stability was not an issue until we began to do preconditioning using polynomial division.

CONTINUOUSLY VARIABLE PEF'S

Inevitably, we get involved with nonstationarity and we become interested in continuously variable PEF's. Again, 1-D theory guides us. We can do regional averaging of c 's and that preserves minimum phase. Likewise, we can independently average across micropatches the numerator and denominator of (2). Many people did this in the old days when filter theory was basically one dimensional.

Polynomial division by nonstationary PEFs made up from an assemblage of stationary ones need not, however, be necessarily stable, as shown by (Rickett, 1999). He demonstrated instability when two stable PEFs alternated at alternate time points. Whether this kind of instability would arise in practice remains to be seen.

ACTION ITEMS

Regretably, I do not recognize any immediately manageable action items. Any Burg-helix PEF estimation method must include a method for polynomial division or it cannot provide fast solutions to the missing data problem. Even with a polynomial division method, we'll still have some thinking to do.

WEB REFERENCE

<http://sepwww.stanford.edu/sep/jon/trash/helburg/> (This paper)
<http://sepwww.stanford.edu/sep/prof/fgdp/c7.ps.gz> (FGDP chapter)
http://sepwww.stanford.edu/sep/prof/fgdp/c7/paper_html/node3.html (html)

ACKNOWLEDGEMENT

I'd like to acknowledge helpful conversations with Sergey Fomel, James Rickett, Francis Muir, and John Burg.

REFERENCES

Claerbout, J. Fundamentals of Geophysical Data Processing:
<http://sepwww.stanford.edu/sep/prof/>, 1976.

Rickett, J., 1999, On non-stationary convolution and inverse convolution: SEP-102, 129-136.

Data alignment with non-stationary shaping filters

James Rickett¹

ABSTRACT

Cross-correlation provides a method of calculating a static shift between two datasets. By cross-correlating patches of data, I can calculate a “warp function” that dynamically aligns the two datasets. By exploiting the link between cross-correlation and shaping filters, I calculate warp functions in a more general way, leveraging the full machinery of geophysical estimation. I compare warp functions, derived by the two methods, for simple one and two-dimensional applications. For the one-dimensional well-tie example, shaping filters gave significantly improved results; however, for the two dimensional residual migration example, the cross-correlation technique gave the better results. I also explain how the helical transform allows the problem of finding a shaping filter to be formulated as an auto-regression.

INTRODUCTION

In both the fields of medical and seismic imaging, automated interpretation of volumetric data is becoming very important. A classical medical imaging problem is how to deform a template image to match an observed image (Kjems et al., 1996). This deformation process is also known as *warping* (Wolberg, 1990), and is the subject of a large literature within the medical imaging community that is reviewed by Toga and Thompson (1998).

Applications of warping, however, are not limited to medical imaging: automated coregistration algorithms may be useful whenever multiple datasets need to be compared directly with one another. In the field of seismic exploration, Grubb and Tura (1997) used a warping algorithm when estimating AVO uncertainties: they migrated a field with multiple equiprobable velocity fields, and colocated the images with cross-correlation derived warp functions. In another seismic application, Rickett and Lumley (1998) included warping as part of a time-lapse reservoir monitoring cross-equalization flow, specifically to address the effects of migrations with different velocity fields. They also found a link between statistically derived warp functions and deterministic residual migration (Rothman et al., 1985) or velocity continuation (Fomel, 1997a) operators.

¹email: james@sep.stanford.edu

Warping: a two step process

Warping is nothing more than resampling of an image in a stretched coordinate system. I split the warping operation into two stages: firstly, the determination of the new coordinate system, and secondly, the spatial and/or temporal resampling.

The second step, the resampling, is just an interpolation operation. It is no more complex than applying normal moveout (NMO), for example. However, as with other resampling operators, the choice of interpolation (nearest-neighbour, linear, band-limited etc.) may affect the quality of results.

The more challenging problem is how do we determine the new coordinate system? Or equivalently, how do we determine the warp function itself?

Calculation of the warp function

The previous authors (Grubb and Tura, 1997; Rickett and Lumley, 1998) determined the warp function by calculating local cross-correlation functions at certain node points within the image space. The warp function at those node points is then taken to be the lag associated with the maxima of the cross-correlelograms. Node values may then be smoothed and interpolated to fill the image volume.

Unfortunately because seismic data is band-limited, cross-correlelograms have multiple local maxima, and in noisy data situations, the maxima may have similar amplitude. Simply picking maxima is, therefore, prone to cycle-skipping problems. The problem of cycle-skipping is fundamental to the process of automated (and even manual) seismic interpretation. Rickett and Lumley (1998) addressed it by heavy smoothing and median filtering of the warp functions before the image resampling itself.

In this paper, I exploit the link between shaping filters and cross-correlation functions to incorporate the smoothing within the matching process. I pick shaping filter maxima as opposed to cross-correlelogram maxima, reducing the need for ad hoc smoothing after the fact.

THEORY

Cross-correlation and shaping filters

A simple numerical way to find a static shift between two traces is to find the maximum of their cross-correlation function. The relative shift is the corresponding cross-correlation lag. Shaping filters are closely related to the simple cross-correlation function, and can also be used to measure relative shifts.

The shaping filter designed to match a first dataset, \mathbf{d}_1 , with a second dataset, \mathbf{d}_2 , can be

defined as the filter, \mathbf{a} that minimizes the norm of the objective function,

$$O(\mathbf{a}) = \|\mathbf{a} * \mathbf{d}_1 - \mathbf{d}_2\|, \quad (1)$$

where $*$ denotes convolution. Equation (1) is very general: it implies nothing about either the choice of norm, or the dimensionality of \mathbf{d}_1 , \mathbf{d}_2 or the filter \mathbf{a} .

The classical discrete solution (Robinson and Treitel, 1980) to equation (1), which minimizes $O(\mathbf{a})$ in the L_2 sense, can be written as

$$\mathbf{a} = (\mathbf{D}_1^T \mathbf{D}_1)^{-1} \mathbf{D}_1^T \mathbf{d}_2. \quad (2)$$

In this paper, I will use the convention that a bold upper case letter represents the operator that describes convolution with the filter represented by the corresponding lower case letter. For example, \mathbf{D}_1 represents the matrix which describes convolution with the dataset, \mathbf{d}_1 , and \mathbf{D}_2 describes the matrix which represents convolution with \mathbf{d}_2 . Multi-dimensionality in equation (2) is built into the definition of the convolution matrices.

Equation (2) implies that the optimal shaping filter, \mathbf{a} , is given by the cross-correlation of \mathbf{d}_1 with \mathbf{d}_2 , filtered by the inverse of the auto-correlation of \mathbf{d}_1 . Equation (2) provides an alternative method of computing a cross-correlation function: firstly calculate an L_2 shaping filter to link one dataset with the other; secondly, recolor the filter with the auto-correlation of the first dataset.

It is not immediately clear why we would ever want to do this in practice, since the first step of computing a shaping filter is to compute a cross-correlation. However, shaping filter estimation can leverage the well-developed machinery of geophysical inversion (Claerbout, 1999) in a number of ways; for example, we may include non-stationarity, a different choice of norm, or different types of regularization in an alternative definition of a shaping filter.

The new algorithm for finding a warp function has three steps. First, estimate a non-stationary shaping filter. Second, recolor the shaping filter by convolving it with the autocorrelation of \mathbf{d}_1 . Finally pick the maxima of the recolored shaping filters.

Adaptive shaping filters

The first step is to consider non-stationary shaping filters. Experience with missing data problems (Crawley et al., 1998; Crawley, 1999b) has shown that working with smoothly-varying non-stationary filters often gives better results than working with filters that are stationary within small patches.

With a non-stationary convolution filter, \mathbf{f} , the shaping filter regression equations,

$$\mathbf{A}_1 \mathbf{f} - \mathbf{a}_2 = \mathbf{0}, \quad (3)$$

are massively underdetermined since there is a potentially unique impulse response associated with every point in the dataspace (Rickett, 1999). We need constraints to ensure the filters vary-smoothly in some manner. The simplest regularization scheme involves applying

a generic data-space roughening operator, \mathbf{R} , to the non-stationary filter coefficients. \mathbf{R} can be a simple derivative operator, for example. This leads to the set of equations,

$$\mathbf{A}_1 \mathbf{f} - \mathbf{a}_2 = \mathbf{0} \quad (4)$$

$$\epsilon \mathbf{R} \mathbf{f} = \mathbf{0}. \quad (5)$$

By making the change of variables, $\mathbf{q} = \mathbf{R} \mathbf{f}$ (Fomel, 1997b), we get the following system of equations,

$$\mathbf{A}_1 \mathbf{R}^{-1} \mathbf{q} - \mathbf{a}_2 = \mathbf{0} \quad (6)$$

$$\epsilon \mathbf{q} = \mathbf{0}. \quad (7)$$

Equations (6) and (7) describe a preconditioned linear system of equations, the solution to which converges rapidly under an iterative conjugate-gradients solver. In practice, I set $\epsilon = 0$, and keep the filters smooth by restricting the number of iterations (Crawley, 1999a).

Shaping filters on a helix

In a helical coordinate system (Claerbout, 1997), calculation of shaping filters can be formulated as an autoregression. If we concatenate the two datasets being matched, and ensure that filter lags span the two datasets, then the shaping filter is identical to the prediction-filter that is used to predict the second dataset from the first. This convenient observation allows reuse of both codes and concepts.

APPLICATIONS

I demonstrate the difference in results obtained by cross-correlations and non-stationary shaping filters on two simple examples: firstly, a 1-D well-tie problem, and secondly a 2-D statistical residual migration. Both methods extend naturally to higher dimensions as well; however, while the cost of the cross-correlation method is proportional to the number of filter coefficients, N_f , the cost of the shaping filter method is proportional to N_f^2 , since the number of iterations depends on the number of filter coefficients.

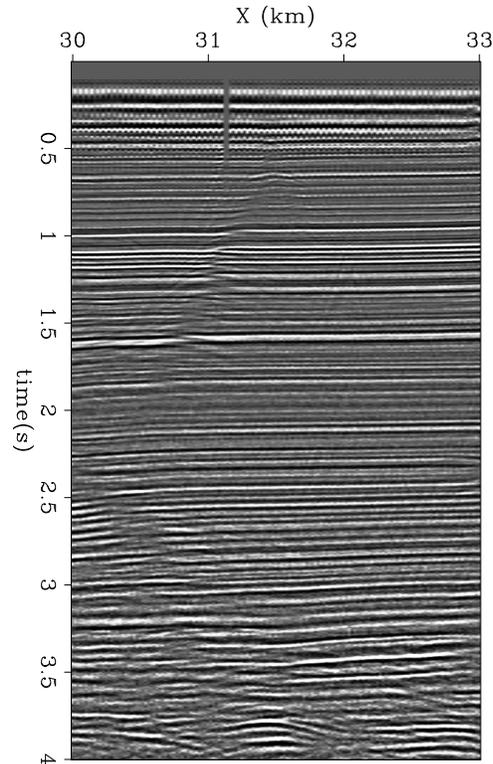
One-dimensional well-ties

Figure 1 shows a seismic section from which I drew two hypothetical sonic logs. The task was then to match them to each other. The fault that cuts through the section and the poor data quality in the lower part cause problems from matching algorithms.

Figures 2 and 3 compare the results of the two warping algorithms. The shaping filters themselves change more smoothly as a function of time than the local cross-correlelograms. The warped difference (trace 3 in both figures) also contains less energy in Figure 3 than Figure 2 indicating the shaping filters have found a better match. Clip levels for the wiggle displays are the same for both Figures 2 and 3.

Figure 1: Seismic section from which two hypothetical sonic logs were drawn. \mathbf{d}_1 was drawn at $X=31$ km and \mathbf{d}_2 was drawn at $X=32$ km.

`james2-data` [ER]



Two-dimensional residual migration

The two panels in Figure 4 show the same 2-D section from a 3-D depth migration of the Elf L7D dataset (Vaillant, 1999), after residual migration with two different residual velocity ratios (Sava, 2000) and conversion to traveltimes depth. I converted to traveltimes depth to avoid remove systematic gross kinematic changes in depth, while leaving changes in kinematics due to imaging differences.

Figure 5 compares the raw difference between the two panels in Figure 4 with the difference after warping with the cross-correlation algorithm, and the shaping filter algorithm. Comparing panel (a) and panel (b) shows the energy in the cross-correlation difference section has decreased significantly, and the algorithm has aligned the main events well. The RMS amplitude decreased a factor of 0.6. On the other hand, comparing panels (a) and (c) shows the shaping filters have not nearly been as successful: although the flanks of the salt have been well aligned, other parts of the section are less well aligned than before warping. The RMS amplitude in panel (c) has only been reduced by a factor of 0.8.

Although I tried a range of parameters, the shaping filter results could probably be improved by tweaking them further. There are many possible options with regard choice of roughening filter and number of iterations, and it is difficult to find the best set.

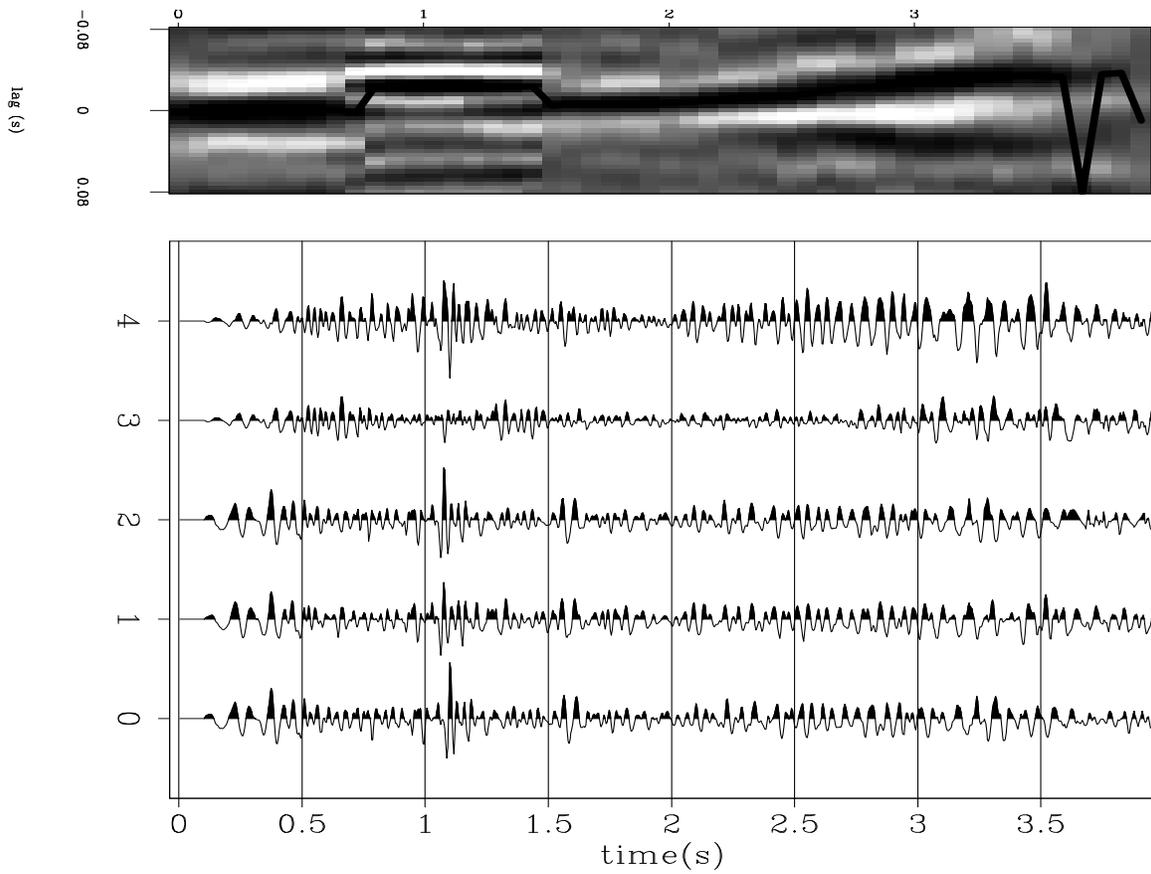


Figure 2: Cross-correlation results. The top panel shows cross-correlelograms and picked maxima. The bottom panel shows \mathbf{d}_1 (0), \mathbf{d}_2 (1), warped \mathbf{d}_1 (2), $\mathbf{d}_2 - \text{warped } \mathbf{d}_1$ (3), and $\mathbf{d}_2 - \mathbf{d}_1$ (4). `james2-xcorr1` [ER]

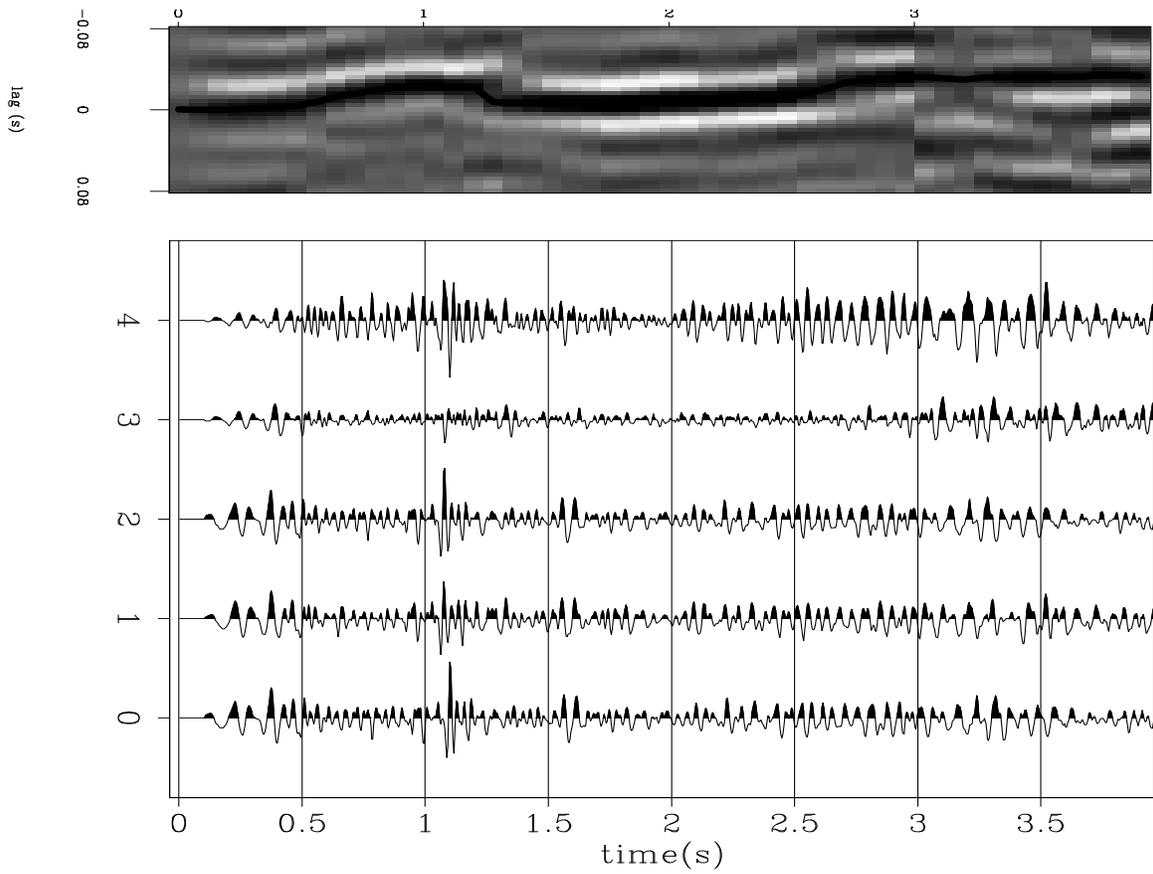


Figure 3: Non-stationary shaping filter results. The top panel shows cross-correlelograms and picked maxima. The bottom panel shows \mathbf{d}_1 (0), \mathbf{d}_2 (1), warped \mathbf{d}_1 (2), $\mathbf{d}_2 - \text{warped } \mathbf{d}_1$ (3), and $\mathbf{d}_2 - \mathbf{d}_1$ (4). `james2-xcorr2` [ER]

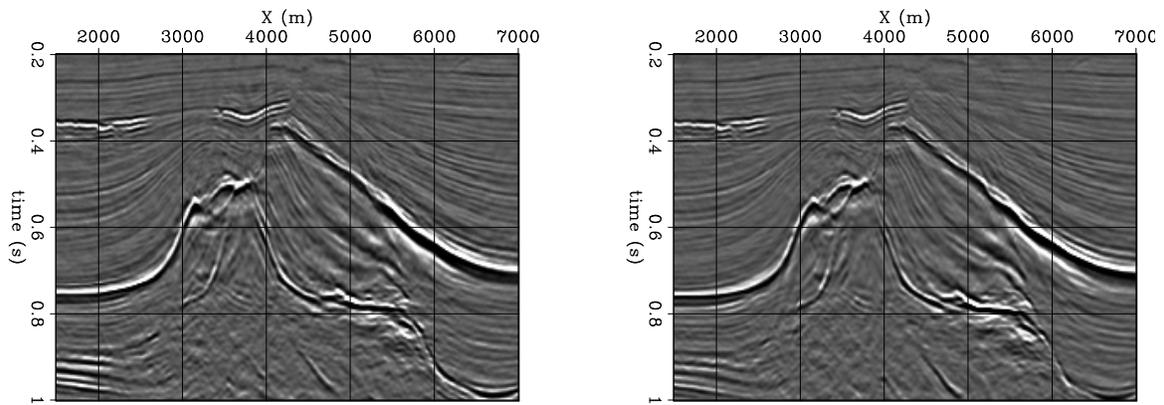
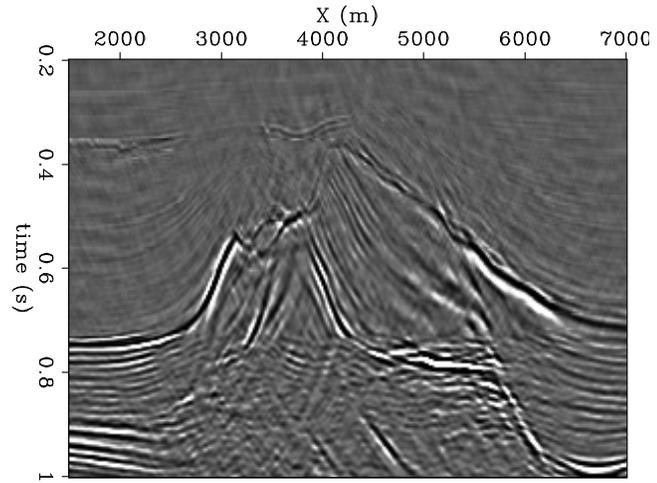
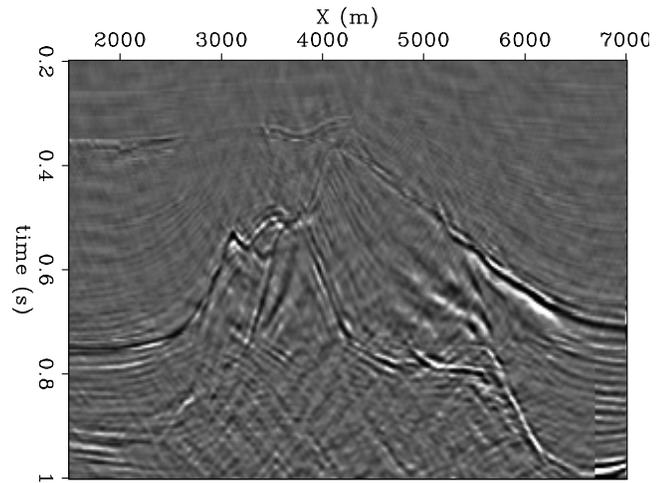


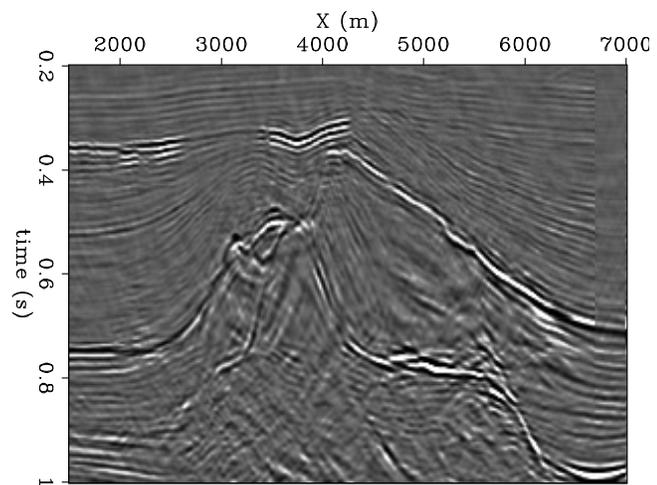
Figure 4: Seismic sections extracted from the Elf L7D dataset after residual migration with $v_{\text{ratio}} = 0.96$ and $v_{\text{ratio}} = 1.02$ and conversion from depth to travelttime depth. `james2-elfpanels` [ER]



(a)



(b)



(c)

Figure 5: (a) Raw difference between the two panels in Figure 4. (b) The difference after warping with the cross-correlation algorithm. (c) The difference after warping with the shaping filter algorithm.

`james2-diff` [CR]

FURTHER WORK

Derivation of a warp function is fundamentally a non-linear process, and I am never going to be able to escape that fact. However, there are tractable and non-tractable non-linear problems that we have some experience with. By formulating the problem of finding a warp function in the framework of geophysical estimation theory, I have opened up many possible avenues of exploration.

A big problem is the presence of secondary maxima that confuse the picking operation; ideally we would like to pick from functions with unique maxima. We may be able to achieve this goal by imposing a minimum-entropy condition (Thorson, 1984) on the shaping filter regularization (model-space residual). In practice, however, this may have convergence problems, and we may be better off with an L_1 norm on the model-space residual (Nichols, 1994), or even the Huber norm (Guitton, 2000).

CONCLUSIONS

Shaping filters are closely related to cross-correlelograms, and therefore can be used to calculate kinematic misalignments between two similar datasets. I demonstrate the use of shaping filters to calculate a dynamic “warp” function that maps one dataset to another. Shaping filters can leverage the power of geophysical estimation theory, which potentially may help avoid problems associated with noisy data to provide improved estimates of multi-dimensional warp functions.

I compared results of warping with a function derived from shaping filters with results from a warp function derived from cross-correlelograms. For the one-dimensional well-tie example, the shaping filters gave encouraging results; however, for the two-dimensional example the cross-correlation technique gave better results.

ACKNOWLEDGEMENTS

Thanks to Paul Sava for providing the residual migrations for the 2-D example.

REFERENCES

- Claerbout, J., 1997, Preconditioning and scaling: SEP-94, 137–140.
- Claerbout, J., 1999, Geophysical estimation by example: Environmental soundings image enhancement: Stanford Exploration Project, <http://sepwww.stanford.edu/sep/prof/>.
- Crawley, S., Clapp, R., and Claerbout, J., 1998, Decon and interpolation with nonstationary filters: SEP-97, 183–192.

- Crawley, S., 1999a, Interpolating missing data: convergence and accuracy, t and f : SEP-**102**, 77–90.
- Crawley, S., 1999b, Interpolation with smoothly nonstationary prediction-error filters: SEP-**100**, 181–196.
- Fomel, S., 1997a, Velocity continuation and the anatomy of prestack residual migration: 67th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1762–1765.
- Fomel, S., 1997b, On model-space and data-space regularization: A tutorial: SEP-**94**, 141–164.
- Grubb, H., and Tura, A., 1997, Interpreting uncertainty measures for AVO migration/inversion: 67th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 210–213.
- Guitton, A., 2000, Implementation of a nonlinear solver for minimizing the Huber norm: SEP-**103**, 281–289.
- Kjems, U., Hansen, L. K., and Chen, C. T., 1996, A non-linear 3D brain co-registration method Proceedings of the Interdisciplinary Inversion Workshop 4. Methodology and Applications in Geophysics, Astronomy, Geodesy and Physics, Technical University of Denmark.
- Nichols, D., 1994, Velocity-stack inversion using L_p norms: SEP-**82**, 1–16.
- Rickett, J., and Lumley, D., 1998, A cross-equalization processing flow for off-the-shelf 4-D seismic data: 68th Ann. Internat. Meeting, Soc. Expl. Geophys., Expanded Abstracts, 16–19.
- Rickett, J., 1999, On non-stationary convolution and inverse convolution: SEP-**102**, 129–136.
- Robinson, E. A., and Treitel, S., 1980, Geophysical signal analysis: Prentice-Hall.
- Rothman, D. H., Levin, S. A., and Rocca, F., 1985, Residual migration - applications and limitations: Geophysics, **50**, 110–126.
- Sava, P., 2000, Variable-velocity prestack Stolt residual migration with application to a North Sea dataset: SEP-**103**, 147–157.
- Thorson, J. R., 1984, Velocity stack and slant stack inversion methods: SEP-**39**.
- Toga, A. W., and Thompson, P. M., 1998, Brain warping:, *in* Brain Warping Academic Press, 1–26.
- Vaillant, L., 1999, Elf L7d North Sea 3-D dataset: SEP Data Library.
- Wolberg, G., 1990, Digital image warping: IEEE Computer Society.

Short Note

Random lines in a plane

Jon Claerbout¹

INTRODUCTION

Locally, seismic data is a superposition of plane waves. The statistical properties of such superpositions are relevant to geophysical estimation and they are not entirely obvious.

Clearly, a planar wave can be constructed from a planar distribution of point sources. Contrariwise, a point source can be constructed from a superposition of plane waves going in all directions. We can represent a random wave source either as a superposition of points or as a superposition of plane waves. Here is the question:

Given a superposition of infinitely many impulsive plane waves of random amplitudes and orientations, what is their spectrum?

If you said the spectrum is white, you guessed wrong. Figure 1 shows that it does not even look white. It is lower frequency than white for good reason. Mathematically independent variables are not necessarily statistically independent variables.

RESOLUTION OF THE PARADOX

If we throw impulse functions randomly onto a plane, the power spectrum of the plane is the power spectrum of impulse functions, namely white.

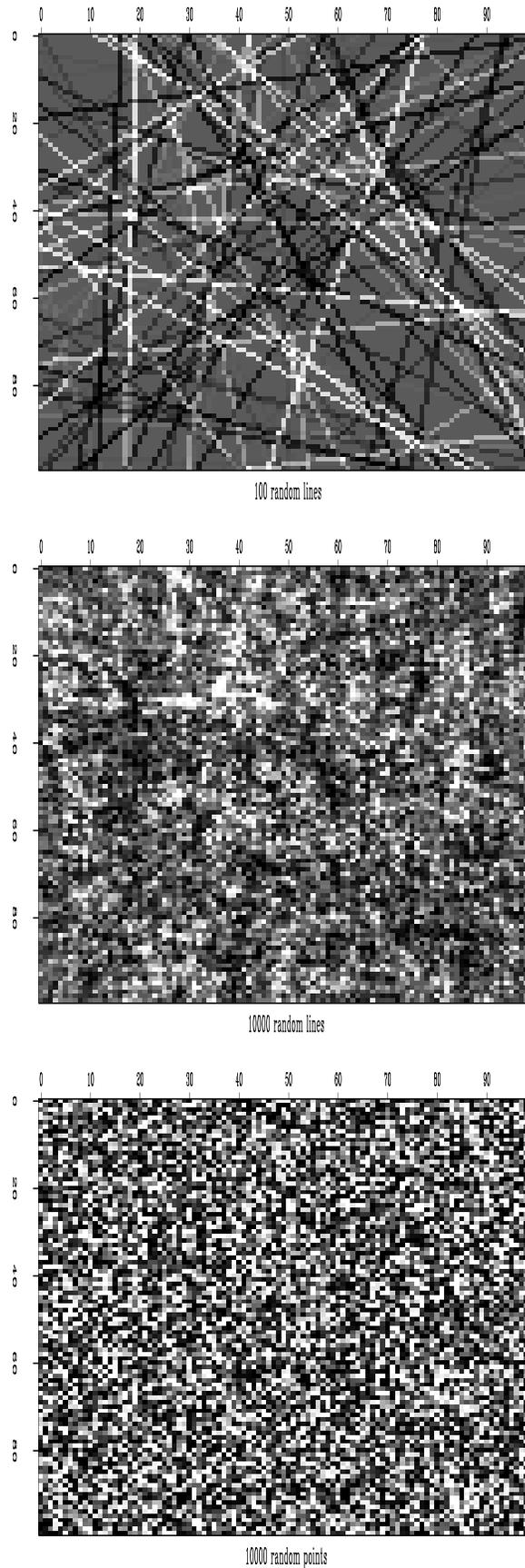
Think of a 2-D Gaussian whose contour of half-amplitude describes an ellipse of great eccentricity. In the limit of large eccentricity, this Gaussian could be one of the lines that we sprinkle on the plane with random amplitudes and orientations. The spatial spectrum of such an eccentric Gaussian must be lower than that of a symmetrical point Gaussian because the spectrum along the long axis of the ellipsoid is concentrated at very low frequency.

Consider a single delta function along a line with an arbitrary slope and location in a plane. The autocorrelation of this dipping line is another dipping line with the same slope, but passing through the origin at zero lag. The polarity of the impulse function is lost in the autocorrelation; in the autocorrelation space, the amplitude of the dipping line is positive.

¹email: claerbout@stanford.edu

Figure 1: Top shows a superposition of 100 randomly positioned lines. Middle shows a superposition of 10,000 such lines. Bottom shows a superposition of 10,000 random point values. The bottom panel shows the most high frequency. Its spectrum is theoretically white. This paper claims that the middle panel is more representative of natural noises.

`jon3-randline` [ER,M]



Now consider a superposition of many dipping lines on the plane. Its autocorrelation is the sum of the autocorrelations of individual lines. The autocorrelation of any individual line is a line of the same slope that is translated to pass through the origin. [The 2-D autocorrelation is not shown in the graphics here. You'll need to understand it from the words here. Sorry.] The autocorrelation is a superposition of lines of various slopes all passing through the origin, all having positive amplitude. This function would resemble a positive impulse function at the origin (and hence suggest a white spectrum). The function is actually not an impulse function, but, as we'll see, it is the pole $1/r$.

Consider an integral on a circular path around the origin. The circle crosses each line exactly twice. Thus the integral on this circular path is independent of the radius of the circle. Hence the average amplitude on the circumference is inverse with the circumference to keep the integral constant. Thus the autocorrelation function is the pole $1/r$.

FOURIER TRANSFORM OF $1/R$

We would like to know the 2-D Fourier transform of $1/r$. Everywhere I found tables of 1-D Fourier transforms but only one place did I find a table that included this 2-D Fourier transform. It was at <http://www.ph.tn.tudelft.nl/Courses/FIP/noframes/fip-Statisti.html>

Sergey Fomel showed me how to work it out: Express the FT in radial coordinates:

$$\text{FT}\left(\frac{1}{r}\right) = \int \int \exp[ik_x r \cos \theta + ik_y r \sin \theta] \frac{1}{r} r dr d\theta \quad (1)$$

$$\text{FT}\left(\frac{1}{r}\right) = \int \delta[k_x \cos \theta + k_y \sin \theta] d\theta \quad (2)$$

To evaluate the integral, we use the fact that $\int \delta(f(x))dx = 1/|f'(x_0)|$ where x_0 is defined by $f(x) = 0$ and the definition $\theta_0 = \arctan(-k_x/k_y)$.

$$\text{FT}\left(\frac{1}{r}\right) = \frac{1}{|-k_x \sin \theta_0 + k_y \cos \theta_0|} \quad (3)$$

$$\text{FT}\left(\frac{1}{r}\right) = \frac{1}{\sqrt{k_x^2 + k_y^2}} = \frac{1}{k_r} \quad (4)$$

UTILITY OF THIS RESULT

At present we are accustomed to estimating statistical properties of seismic data by computing a 2-D prediction-error filter. This filter is needed to interpolate and extrapolate missing values.

Knowing that the prior spectral estimate is not a constant but instead is $1/k_r$ suggests a procedure that is more efficient statistically: By more efficient, I mean that a simpler model should fit the data, a model with fewer adjustable parameters.

INTERPOLATION

We'll need to know a wavelet in the time and space domain whose amplitude spectrum is $\sqrt{k_r}$ (so its power spectrum is k_r). Do not mistake this for the the helix derivative (Claerbout, 1998) whose power spectrum is k_r^2 . What we need to use here is the square root of the helix derivative. Let the (unknown) wavelet with amplitude spectrum $\sqrt{k_r}$ be known as G .

1. Apply G to the data. The prior spectrum of the modified data is now white.
2. Estimate the PEF of the modified data.
3. The interpolation filter for the original data is now G times the PEF of the modified data.

Why is this more efficient? The important point is that the PEF should estimate the minimal practical number of freely adjustable parameters. If G is a function that is lengthy in time or space, then the PEF does not need to be.

How important is this extra statistical efficiency? I don't know.

WHAT ARE THE NEXT STEPS?

- Compute $\sqrt{k_r}$ in physical space and look at it. How best to do this? How best to package the software?
- Invent a synthetic data test.
- Think about how it might impact Sean (or Sergey). Which example of Sean's would be worth redoing?
- Extend this idea to 3-D (lettuce versus noodles).
- Matt Schwab and I were frequently disappointed in the performance of local PEFs for the task of visualizing data. This might explain it. Which example of his might be worth redoing?

REFERENCES

Claerbout, 1998, Multidimensional recursive filters via a helix: *Geophysics*, **63**, 1532–1541.

Short Note

Test case for PEF estimation with sparse data

Jon Claerbout¹

Are PEFs going to be a useful tool on very sparse data sets? GEE disappoints. It throws away regression equations that depend on missing data. With sparse data, we have none left.

We need a test case. I offer one here. A test case with some realistic aspects. We should be able to get perfect results while nobody else does. And they'll all have to admit it :-)

THE TEST CASE

We begin with a rough one-dimensional function. A random walk would be nice, the integral of random numbers (possibly coin flips). Call it $r(x)$. Actually, I'd like a random walk that crosses the zero axis a couple times. We could try several seeds until we find an "attractive" one. Maybe leaky integrate random numbers.

Next, flex a piece of paper so that along the x -axis it matches $r(x)$. Now any line parallel to the x -axis should match $r(x)$. Let us tilt this on the y -axis, so our altitude function is $h(x, y) = r(x) + y$. Notice that we have a function whose second y -derivative vanishes everywhere. So δ_{yy} is a PEF for it. Its Gaussian curvature $h_{xx}h_{yy} - h_{xy}^2$ vanishes. Now, to confuse people, we rotate it 45° . Thus, I propose the test function $h(x, y) = r(x + y) + (x - y)$. Its perfect PEF is

$$\begin{array}{ccc} . & 0 & 0 \\ . & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{array}$$

and we need to show that we can find it.

What will we give for data? I offer a track of dense data along the x axis (like a well log or seismogram). Already you can see with this and with the PEF, and one more data point off the track, the data everywhere is known. Well, almost everywhere. Actually, there is a shadow. Now we sprinkle a dozen data values around the plane. We might choose those points aliased on our rough function in such a way that everyone but us would be completely confused.

¹email: claerbout@stanford.edu

HOW ARE WE GOING TO GET THE PEF?

How are we going to get the PEF? I propose we invoke stationarity and scale invariance. Notice that we have an interesting kind of scale invariance here. We can expand the correct PEF and it is still a correct PEF. (It is interesting to notice that an apparent scale or resonant spectrum visible along the observation track does not prevent the 2-D function from being scale invariant).

We have a non-linear optimization problem to solve. You know it, $0 \approx A(Z)P(Z)$ where $a_0 = 1$ and some of the p_i are known. It is non-linear. This problem is linearized in (Claerbout, 1992) and (Claerbout, 2000) but I don't trust an ignorant descent. I've had some disappointments doing that (but that was before I learned about scale invariance). How would you try?

REFERENCES

- Claerbout, J. F., 1992, *Earth Soundings Analysis: Processing versus Inversion*: Blackwell Scientific Publications.
- Claerbout, J. F., 2000, *Geophysical Estimation by Example: Environmental soundings image enhancement*: <http://sepwww.stanford.edu/sep/prof/>.

Transformation of seismic velocity data to extract porosity and saturation values for rocks

James G. Berryman, Patricia A. Berge, and Brian P. Bonner

ABSTRACT

For wave propagation at low frequencies in a porous medium, the Gassmann-Domenico relations are well-established for homogeneous partial saturation by a liquid. They provide the correct relations for seismic velocities in terms of constituent bulk and shear moduli, solid and fluid densities, porosity and saturation. It has not been possible, however, to invert these relations easily to determine porosity and saturation when the seismic velocities are known. Also, the state (or distribution) of saturation, *i.e.*, whether or not liquid and gas are homogeneously mixed in the pore space, is another important variable for reservoir evaluation. A reliable ability to determine the state of saturation from velocity data continues to be problematic. We show how transforming compressional and shear wave velocity data to the $(\rho/\lambda, \mu/\lambda)$ -plane (where λ and μ are the Lamé parameters and ρ is the total density) results in a set of quasi-orthogonal coordinates for porosity and liquid saturation that greatly aids in the interpretation of seismic data for the physical parameters of most interest. A second transformation of the same data then permits isolation of the liquid saturation value, and also provides some direct information about the state of saturation. By thus replotting the data in the $(\lambda/\mu, \rho/\mu)$ -plane, inferences can be made concerning the degree of patchy (inhomogeneous) versus homogeneous saturation that is present in the region of the medium sampled by the data. Our examples include igneous and sedimentary rocks, as well as man-made porous materials. These results have potential applications in various areas of interest, including petroleum exploration and reservoir characterization, geothermal resource evaluation, environmental restoration monitoring, and geotechnical site characterization.

INTRODUCTION

In a variety of applied problems, it is important to determine the state of saturation of a porous medium from acoustic or seismic measurements. In the oil and gas industry, it is common to use amplitude-versus-offset (AVO) processing of seismic reflection data to reach conclusions about the presence of gas, oil, and their relative abundances on the opposite sides of a reflecting interface underground [*e.g.*, Castagna and Backus (1993)]. For environmental applications, we can expect to be working in the near surface where sensor geometries other than surface reflection surveys become practical. For example, when boreholes are present, it is possible to do crosswell seismic tomography, or borehole sonic logging to determine ve-

locities [*e.g.*, Harris *et al.* (1995)]. For AVO processing the data obtained are the seismic impedances ρv_p and ρv_s (where ρ is the density, and v_p, v_s are the seismic compressional and shear wave velocities, respectively), which arise naturally in reflectance measurements. (In this paper, we will use the term “velocities” to refer to measured velocities at seismic, sonic, or ultrasonic frequencies, unless otherwise specified.) However, for crosswell applications, we are more likely to have simply velocity data, *i.e.*, v_p and v_s themselves without density information. For well-logging applications, separate measurements of the velocities as well as density are possible. Although a great deal of effort has been expended on AVO analysis, relatively little has been done to invert the simple velocity data for porosity and saturation. It is our purpose to present one method that shows promise for using velocity data to obtain porosity and saturation estimates. The key physical idea used here is the fact that the Lamé parameter λ and the density ρ are the two parameters containing information about saturation, while both of these together with shear modulus μ contain information about porosity (λ and μ are defined in the next section). These facts are well-known from earlier work of Gassmann (1951), Domenico (1974), and many others. (It is well-established that even though the Gassmann-Domenico relations are derived for the static case, they have been found to describe behavior measured in the field at sonic and seismic frequencies, and, in some cases, even in laboratory ultrasonic experiments.) The same facts are used explicitly in AVO analysis (Castagna and Backus, 1993; Ostrander, 1984; Castagna *et al.*, 1985; Foster *et al.*, 1997), but in ways that are significantly different from those to be described here. A major point of departure is that the present work allows direct information to be obtained about, not only the level of the saturation, but also concerning the state of saturation, *i.e.*, whether the liquid and gas present are mixed homogeneously, or are instead physically separated and therefore in a state of patchy saturation (Berryman *et al.*, 1988; Endres and Knight, 1989; Knight and Nolen-Hoeksema, 1990; Mavko and Nolen-Hoeksema, 1994; Dvorkin and Nur, 1998; Cadoret *et al.*, 1998). Another advantage is that this method uses velocity rather than amplitude information, and therefore may have less uncertainty and may also require less data processing for some types of field experiments.

One of the main points of the analysis to be presented is the purposeful avoidance of the well-known complications that arise at high frequencies, due in large part to velocity dispersion and attenuation (Biot, 1956a,b; Biot, 1962; O’Connell and Budiansky, 1977; Mavko and Nur, 1978; Berryman, 1981; McCann and McCann, 1985; Johnson *et al.*, 1987; Norris, 1993; Best and McCann, 1995). Our point of view is that seismic data (as well as most sonic, and some ultrasonic data) do not suffer from contamination by the frequency-dependent effects to the same degree typically seen for high frequency laboratory measurements. By restricting our range of frequencies to those most useful in the field, we anticipate a significant simplification of the analysis and therefore an improvement in our ability to provide both simple and robust interpretations of field data. In the Discussion section, we also provide a means of identifying data in need of correction for dispersion effects.

We introduce the basic physical ideas in the next section. Then we present two new methods of displaying the velocity data. One method is used to sort data points into sets that have similar physical attributes, such as porosity. Then, the second method is used to identify both the level of saturation and the type of saturation, whether homogeneous, patchy, or a combina-

tion of the two. We show a subset of the large set of data we have examined that confirms these conclusions empirically. We then provide some discussion of the results and what we foresee as possible future applications of the ideas. Finally, we summarize the accomplishments of the paper in the concluding section.

ELASTIC AND POROELASTIC WAVE PROPAGATION

For isotropic elastic materials there are two bulk elastic wave speeds (Aki and Richards, 1980), compressional $v_p = \sqrt{(\lambda + 2\mu)/\rho}$ and shear $v_s = \sqrt{\mu/\rho}$. Here ρ is the overall density, and the Lamé parameters λ and μ are the constants that appear in Hooke's law relating stress to strain in an isotropic material. The constant μ gives the dependence of shear stress on shear strain in the same direction. The constant λ gives the dependence of compressional or tensional stress on extensional or dilatational strains in orthogonal directions. For a porous system with porosity ϕ (void volume fraction) in the range $0 < \phi < 1$, the overall density of the rock or sediment is just the volume weighted density given by

$$\rho = (1 - \phi)\rho_s + \phi[S\rho_l + (1 - S)\rho_g], \quad (1)$$

where ρ_s , ρ_l , ρ_g are the densities of the constituent solid, liquid and gas, respectively. S is the liquid saturation, *i.e.*, the fraction of liquid-filled void space in the range $0 \leq S \leq 1$ [see Domenico (1974)]. When liquid and gas are distributed uniformly in all pores and cracks, Gassmann's equations say that, for quasistatic isotropic elasticity and low frequency wave propagation, the shear modulus μ will be mechanically independent of the properties of any fluids present in the pores, while the overall bulk modulus K ($\equiv \lambda + \frac{2}{3}\mu$) of the rock or sediment including the fluid depends in a known way on porosity and elastic properties of the fluid and dry rock or sediment (Gassmann, 1951; Berryman, 1999). Thus, in the Gassmann model, the Lamé parameter λ is elastically *dependent* on fluid properties, while μ is not. The density ρ also depends on saturation, as shown in equation (1). At low liquid saturations, the bulk modulus of the fluid mixture is dominated by the gas, and therefore the effect of the liquid on λ is negligible until the porous medium approaches full saturation. This means that both velocities v_p and v_s will decrease with increasing fluid saturation (Domenico, 1974) due to the "density effect," wherein the only quantity changing is the density, which increases in the denominators of both v_p^2 and v_s^2 . As the medium approaches full saturation, the shear velocity continues its downward trend, while the compressional velocity suddenly (over a very narrow range of saturation values) shoots up to its full saturation value. A well-known example of this behavior was provided by Murphy (1984). Figure 1 shows how plots of these data for sandstones will appear in several choices of display, with Figure 1(a) being one of the more common choices. This is the expected (ideal Gassmann-Domenico) behavior of partially saturated porous media. The Gassmann-Domenico relations hold for frequencies low enough (sonic and below) that the solid frame and fluid will move in phase, in response to applied stress or displacement. The fluid pressure must be (at least approximately) uniform throughout the porous medium, from which assumption follows the homogeneous saturation requirement.

PREDICTIONS OF THE THEORY AND EXAMPLES

Gassmann-Domenico relations

Gassmann's equations (Gassmann, 1951) for fluid substitution state that

$$K = K_{dr} + \frac{\alpha^2}{(\alpha - \phi)/K_m + \phi/K_f} \quad \text{and} \quad \mu = \mu_{dr}, \quad (2)$$

where K_m is the bulk modulus of the single solid mineral, K_{dr} and μ_{dr} are the bulk and shear moduli of the drained porous frame. The special combination of moduli defined by $\alpha = 1 - K_{dr}/K_m$ is the Biot-Willis parameter (Biot and Willis, 1957). The porosity is ϕ , while K and μ are the effective bulk and shear moduli of the undrained porous medium that is saturated with a fluid mixture having bulk modulus K_f . For partial saturation conditions with homogeneous mixing of liquid and gas, so that all pores contain the same relative proportions of liquid and gas, Domenico (1974) among others shows that

$$1/K_f = S/K_l + (1 - S)/K_g. \quad (3)$$

The saturation level of liquid is S lying in the range $0 \leq S \leq 1$. The bulk moduli are: K_l for the liquid, and K_g for the gas. When S is small, (3) shows that $K_f \simeq K_g$, since $K_g \ll K_l$. As $S \rightarrow 1$, K_f remains close to K_g until S closely approaches unity. Then, K_f changes rapidly (over a small range of saturations) from K_g to K_l . (Note that the value of K_l may be several orders of magnitude larger than K_g , as in the case of water and air — 2.25 GPa and 1.45×10^{-4} GPa, respectively.)

Since μ has no mechanical dependence on the fluid saturation, it is clear that all the fluid dependence of $K = \lambda + \frac{2}{3}\mu$ in (2) resides within the Lamé parameter λ . Other recent work (Berryman *et al.*, 1999) on layered elastic media indicates that λ should be considered as an important independent variable for analysis of wave velocities and Gassmann's results provide some confirmation of this deduction (and furthermore provided a great deal of the motivation for the present line of research). The parameters $K (= \lambda + \frac{2}{3}\mu)$ and $K_{dr} (= \lambda_{dr} + \frac{2}{3}\mu_{dr})$ can be replaced in (2) by λ and λ_{dr} without changing the validity of the equation. Thus, like K , for increasing saturation values, λ will be almost constant until the porous medium closely approaches full saturation.

Now, the first problem that arises with field data is that we usually do not know the reason why data collected at two different locations in the earth differ. It could be that the differences are all due to the saturation differences we are concentrating on in this paper. Or it could be that they are due entirely or only partly to differences in the porous solids that contain the fluids. In fact, solid differences easily can mask any fluid differences because the range of detectable solid mechanical behavior is so much greater than that of the fluids (especially when fractures are present).

It is essential to remove such differences due to solid heterogeneity. A related issue concerns differences arising due to porosity changes throughout a system of otherwise homogeneous solids. One way of doing this would be to sort our data into sets having similar porous

solid matrix. For simplicity and because of the types of laboratory data sets available, we will use porosity here as our material discriminant.

TABLE 1. Monotonicity properties of the Lamé parameters λ and μ and the density ρ as the porosity ϕ and liquid saturation S vary.

	Lamé λ	Lamé μ	Density ρ
$\Delta\phi$	$\frac{\partial\lambda}{\partial\phi} _S = ?$	$\frac{\partial\mu}{\partial\phi} _S < 0$	$\frac{\partial\rho}{\partial\phi} _S < 0$
ΔS	$\frac{\partial\lambda}{\partial S} _\phi > 0$ (or $\simeq 0$)	$\frac{\partial\mu}{\partial S} _\phi = 0$	$\frac{\partial\rho}{\partial S} _\phi > 0$

Considering our three main parameters, λ , μ , and ρ , we see that all three depend on porosity, but only λ and ρ depend on saturation. Using formulas (1)-(3), we can take partial derivatives of each of these expressions first with respect to ϕ while holding S constant, and then with respect to S while holding ϕ constant. For now, we are only interested in trends rather than the exact values, and these are displayed in Table 1. The trend for $\partial\lambda/\partial S|_\phi > 0$ requires the additional reminder that, although this term is always positive, its value is often so small that it may be treated as zero except in the small range of values close to $S = 1$. Also, using Hashin-Shtrikman bounds (Hashin and Shtrikman, 1962) as a guide, it turns out that it is not possible to make a general statement about the sign of $\partial\lambda/\partial\phi|_S$, since the result depends on the particular material constants. (Related differences of sign are also observed in the data we show later in this paper; thus, this ambiguity is definitely real and observable.)

Assuming that the primary variables are λ , μ , and ρ (further justification of this choice of primary variables is provided later in the paper), then the two pieces of velocity data we have can be used to construct the following three ratios:

$$\frac{\mu}{\lambda} = \frac{v_s^2}{v_p^2 - 2v_s^2}, \quad (4)$$

$$\frac{\rho}{\lambda} = \frac{1}{v_p^2 - 2v_s^2}, \quad (5)$$

and

$$\frac{\rho}{\mu} = \frac{1}{v_s^2}. \quad (6)$$

We will consider first of all what happens to these ratios for homogeneous mixing of fluids, and then consider the simpler case of ideal patchy saturation, where some pores in the partially saturated medium are completely filled with liquid and others are completely dry (or filled with gas).

Homogeneous saturation

For homogeneous saturation, as S varies while porosity remains fixed, the ratio μ/λ does not change significantly until $S \rightarrow 1$. At that point, λ increases dramatically and μ/λ therefore decreases dramatically. Similarly, as $S \rightarrow 1$, the only changes in ρ/λ over most of the dynamic range of S are in ρ , which increases linearly with S . Then, when ρ is almost at its maximum value, λ increases dramatically, causing the ρ/λ ratio to decrease dramatically. Thus, ρ/λ does not change monotonically with S , but first increases a little and then decreases a lot. These two ratios may be conveniently compared by plotting data from various rocks and man-made porous media examples in the $(\rho/\lambda, \mu/\lambda)$ -plane [see Figure 1(b) and Figure 2]. We see that, when data are collected at approximately equal intervals in S , the low saturation points will all cluster together with nearly constant μ/λ and small increases in ρ/λ , but the final steps as $S \rightarrow 1$ lead to major decreases in both ratios. The resulting plots appear as nearly straight lines in this plane, with drained samples plotting to the upper right and fully saturated samples plotting to the lower left in each of the examples shown in Figure 2. The remaining ratio ρ/μ has the simplest behavior, since ρ increases monotonically in S , and μ does not change. So ρ/μ is a monotonically increasing function of S , and therefore can be considered a useful proxy of the saturation variable S . [Compare Figures 1(c) and 1(d), and see Figure 3.]

Figure 2(a) includes the same sandstone data from Figure 1, along with other sandstone data. Similar data for five limestone samples (Cadoret *et al.*, 1998) are plotted in Figure 2(b). The straight line correlation of the data in the sandstone display is clearly reconfirmed by the limestone data. Numerous other examples of the correlation have been observed. [Fully dry and fully saturated examples are shown here for some of these examples in Figures 2(c) and 2(d), for which partial saturation data were unavailable.] No examples of appropriate data for partially saturated samples have exhibited major deviations from this behavior, although an extensive survey of available data sets has been performed for materials including limestones (Cadoret *et al.*, 1998), sandstones (Murphy, 1984; Knight and Nolen-Hoeksema, 1990), granites (Nur and Simmons, 1969), unconsolidated sands, and some artificial materials such as ceramics and glass beads (Berge *et al.*, 1995). This straight line correlation is a very robust feature of partial saturation data. The mathematical trick that brings about this behavior will now be explained.

Consider the behavior as ϕ increases for fixed S . Two of the parameters (μ and ρ) decrease as ϕ increases, but at different rates, while the third (λ) can have arbitrary variation. [Recall (Bourbié *et al.*, 1987) that rigorous bounds on the parameters are: $0 \leq K < \infty$, $0 \leq \mu < \infty$, $0 < \rho < \infty$, and $-\frac{2}{3}\mu \leq \lambda \leq \infty$.] To understand the behavior on these plots in Figure 1 as ϕ changes, it will prove convenient to consider polar coordinates (r, θ) , defined by

$$r^2 = w^4 \left(\frac{\rho}{\lambda} \right)^2 + \left(\frac{\mu}{\lambda} \right)^2, \quad (7)$$

and

$$\tan \theta = \frac{\mu}{\rho w^2}, \quad (8)$$

where w is an arbitrary scale factor with dimensions of velocity (chosen so that r is a dimensionless radial coordinate for plots like those in Figure 1). Now, if in addition we choose w

to be sufficiently large so that $v_s/w \ll 1$ for typical values of v_s in our data sets, then, using standard perturbation expansions, we have

$$r = \frac{\rho}{\lambda} w^2 \left(1 + \frac{v_s^4}{w^4} \right)^{\frac{1}{2}} \simeq \frac{\rho}{\lambda} w^2 \left(1 + \frac{v_s^4}{2w^4} \right) \quad (9)$$

and

$$\theta = \tan^{-1} \left(\frac{v_s^2}{w^2} \right) \simeq \frac{v_s^2}{w^2}. \quad (10)$$

Thus, the angle θ is well approximated by the ratio in (10), which depends only on the shear velocity v_s . We know the shear velocity is a rather weak function of saturation [e.g., Figure 1(a)], but a much stronger function of porosity [see, for example, Berge *et al.* (1995)]. So we see that the angle in these plots is most strongly correlated with changes in the porosity. In contrast, the radial position r is principally dependent on the ratio ρ/λ , which we have already shown to be a strong function of the saturation S , especially in the region close to full liquid saturation. This analysis shows why the plots in Figures 1(b) and 2 look the way they do and also why we might be inclined to call these quasi-orthogonal (polar) plots of saturation and porosity. Because of the function these plots play in our analysis, we will call them the “data-sorting” plots.

In contrast, the plots in Figure 3 contain information about fluid spatial distribution, as will be discussed at greater length later in this paper. The bulk modulus K_f contains the only S dependence in (2). Thus, for porous materials satisfying Gassmann’s homogeneous fluid conditions and for low enough frequencies, the theory predicts that, if we use velocity data in a two-dimensional plot with one axis being the saturation S and the other being the ratio $\lambda/\mu = (v_p/v_s)^2 - 2$, then the results will lie along an essentially straight (horizontal) line until the saturation reaches $S \simeq 1$ (around 95% or higher), where the curve formed by the data will quickly rise to the value determined by the velocities at full liquid saturation. On such a plot, the drained data appear in the lower left while the fully saturated data appear in the upper right. This behavior is illustrated in Figure 3(a) for Espeil limestone. The behavior of the other plots in Figure 3 will be described below.

Before leaving this discussion of homogeneous saturation, we should note that there is one laboratory saturation technique for which it is known — from direct observations (Cadoret *et al.*, 1998) using x-ray imaging — that very homogeneous liquid-gas mixtures will generally be produced. This method is called “depressurization.” When such data are available (see Figure 3), we expect they will always behave according to the Gassmann-Domenico predictions. In contrast, the more common approach which produces drainage data is less predictable, since the manner and rate of drainage depend strongly on details of particular samples — especially on surface energies that control capillarity and on permeability magnitude and distribution. Thus, the drainage technique can produce homogeneous saturation, or patchy saturation, or anything in between.

Patchy saturation

The preceding analysis centered on homogeneous saturation of porous media. On the other hand, consider a porous medium containing gas and liquid mixed in a heterogeneous manner, so that patches of the medium hold only gas while other patches hold only liquid in the pores. Then, the theory predicts that, depending to some extent on the spatial distribution of the patches, the results will deviate overall from Gassmann's results (although Gassmann's results will hold locally in each individual patch). If we consider the most extreme cases of spatial distribution possible, which are laminated regions of alternating liquid saturation and gas saturation, then the effective bulk modulus will be determined by an average of the two extreme values of (2): $K|_{S=0} = K_{dr}$ and $K|_{S=1}$. Using saturation as the weighting factor, the harmonic mean and the mean are the two well-known extremes of behavior (Hill, 1952). Of these two, the one that differs most from (2) for $0 < S < 1$ is the mean. And, because of K 's linear dependence on both λ and μ , and μ 's independence of S , we therefore have

$$\lambda_{patchy}(S) = (1 - S)\lambda_{dr} + S\lambda|_{S=1}. \quad (11)$$

So, on our plot in the $(\lambda/\mu, \rho/\mu)$ -plane, the results for the mean will again lie along a straight line, but now the line goes directly from the unsaturated value ($S = 0$) to the fully saturated value ($S = 1$) [*e.g.*, Figure 3(e)]. The two straight lines described [the one given by (11) and the horizontal one discussed in the preceding paragraph for saturations up to about 95%] are rigorous results of the theory, and form two sides of a triangle that will contain all data for partially saturated systems, regardless of the type of saturation present. The third side of this triangle provides a rigorous bound on the behavior as full saturation is approached (it just corresponds to the physical requirement that $S \leq 1$, so values with $S > 1$ have no physical significance). In general, heterogeneous fluid distribution can produce points anywhere within the resulting triangle, but not outside the triangle (within normal experimental error).

A brief presentation of some examples (Figure 3) will now follow a reminder of an important and well-known caveat.

Caveats for chemical effects

Some deviations from these conclusions can be expected at the lowest saturations. Chemical effects, which have not been accounted for in the mechanical analysis, can and often do lead to the situation that dry and drained (nearly dry or room dry) samples have somewhat different properties (Bonner *et al.*, 1997). These differences are larger than can be explained by mechanical analyses alone. [For example, see Figures 3(a) and 3(b). Take special note of the three lowest saturation values in these Figures.] We discuss this point at greater length in the Discussion section.

Some illustrative examples

Figure 3 shows three examples of the results obtained with plots in the $(S, \lambda/\mu)$ -plane and in the $(\rho/\mu, \lambda/\mu)$ -plane (using ρ/μ as a proxy for S) for two limestones and one andesite

from laboratory data of Cadoret (1993) and Cadoret *et al.* (1995; 1998). In Figure 3, the true saturation data are used in the Figures on the left and the proxy for saturation (ρ/μ) is used on the right. We therefore call the right hand diagrams “saturation-proxy” plots. Using the interpretations arising for our analysis of Gassmann-Domenico partial-saturation theory, we see that Figures 3(a) and (b) indicate homogeneous mixing of liquid and gas, while Figures 3(e) and (f) indicate extremely patchy mixing, and Figures 3(c) and (d) show an intermediate state of mixing for the drainage data, but more homogeneous mixing (as expected) for the depressurization data. The Espeil limestone was observed to be the most dispersive of all those rocks considered in the data sets of Cadoret (1993) and Carodet *et al.* (1995; 1998). So this case is a very stringent test of the method. In fact, if we were to plot the corresponding data for Espeil limestone at 500 kHz, we would not find such simple and easily interpreted behavior on these plots. Our explanation for this difference between the 500 kHz and 1 kHz results for Espeil limestone is that the dispersion introduces effects not accounted for by the simple Gassmann-Domenico theory, and that there is then no reason to think that our method should work for such high frequencies as 500 kHz. We have found other examples where it does work for frequencies higher than one might expect the method to be valid. The point is that, if we restrict the range of frequencies considered to 1 kHz or less, the method appears to work quite well on most (and perhaps all) samples. [But, at higher frequencies, the solid and fluid can move out of phase and other relations developed by Biot (1956a,b; 1962) and others (O’Connell and Budiansky, 1977; Mavko and Nur, 1978; Berryman, 1981; McCann and McCann, 1985; Johnson *et al.*, 1987; Norris, 1993; Best and McCann, 1995) apply.]

DISCUSSION

Rocks containing more than one mineral

The analysis presented here has been limited for simplicity to the case of single mineral porous rocks. In fact the main parts of the analysis do not change in any significant way if the rock has multiple constituents. The well-known result of Brown and Korringa (1975) states that

$$K = K_{dr} + \frac{\alpha^2}{\alpha/K_s - \phi/K_\phi + \phi/K_f} \quad \text{and} \quad \mu = \mu_{dr}, \quad (12)$$

where K_s is theunjacketed bulk modulus of the composite solid frame, K_ϕ is theunjacketed pore modulus of the composite solid frame, $\alpha = 1 - K_{dr}/K_s$ is the appropriate Biot-Willis (1957) parameter for this situation. The remaining parameters have the same significance as in (2). The functional dependence of K_{sat} on the saturation S is clearly the same in both formulas. If we were trying to infer properties of the solid from these formulas, then of course (12) would be more difficult to interpret. But for our present purposes, we are only trying to infer porosity, saturation values, and saturation state. For these physical parameters, the analysis goes through without change.

On uniqueness of λ -diagrams

Since the possible linear combinations of the elastic bulk and shear moduli (K and μ) are infinite, it is natural to ask why (or if) the choice $\lambda = K - \frac{2}{3}\mu$ is special? Is there perhaps some other combination of these constants that works as well or even better than the choice made here? There are some rather esoteric reasons based on recent work (Berryman *et al.*, 1999) in the analysis of layered anisotropic elastic media that lead us to believe that the choice λ is indeed special, but we will not try to describe these reasons here. Instead we will point out some general features of the two types of plots that make it clear that this choice is generally good, even though others might be equally good or even better in special circumstances. First, in the diagram using the $(\rho/\mu, \lambda/\mu)$ -plane, it is easy to see that *any* plot of data using linear combinations of the form $(\rho/\mu, (\lambda + c\mu)/\mu)$, where c is any real constant, will have precisely the same information and the display will be identical except for a translation of the values along the ordinate by the constant value c . Thus, for example taking $c = \frac{2}{3}$, plots of $(\rho/\mu, K/\mu)$ will have exactly the same interpretational value as those presented here. But, if we now reconsider the data-sorting plot (*e.g.*, Figure 2) for each of these choices, we need to analyze plots of the form $(\rho/(\lambda + c\mu), \mu/(\lambda + c\mu))$. Is there an optimum choice of the parameter c that makes the plots as straight as possible whenever the only variable is the fluid saturation? It is not hard to see that the class of best choices always lies in the middle of the range of values of λ/μ taken by the data. So setting $-c = \frac{1}{2}(\min(\lambda/\mu) + \max(\lambda/\mu))$ will always guarantee that there are very large positive and negative values of $\mu/(\lambda + c\mu)$, and therefore that these data fall reliably (if somewhat approximately) along a straight line. But the minimum value of λ/μ has an absolute minimum of $-\frac{2}{3}$, based on the physical requirement of positivity of K . So $c < \frac{2}{3}$ is a physical requirement, and since $\max \lambda/\mu \simeq +\frac{2}{3}$ is a fairly typical value for porous rocks, it is expected that an optimum value of $c \leq 0$ will generally be obtained using this criterion. Thus, plots based on bulk modulus K instead of λ will not be as effective in producing the quasi-orthogonality of porosity and saturation that we have obtained in the data-sorting style of plotting. We conclude that the choice λ is not unique (some other choices might be as good for special data sets), but it is nevertheless an especially simple choice and is also expected to be quite good for most real data.

Transforming straight lines to straight lines

One important feature concerning connections between the points in the two planes $(\rho/\lambda, \mu/\lambda)$ and $(\rho/\mu, \lambda/\mu)$ is the fact that (with only a few exceptions that will be noted) straight lines in one plane transform into straight lines in the other. For example, points satisfying

$$\frac{\lambda}{\mu} = A + B \frac{\rho}{\mu} \quad (13)$$

in the $(\rho/\mu, \lambda/\mu)$ -plane (where A and B are constant intercept and slope, respectively), then satisfy

$$\frac{\mu}{\lambda} = A^{-1} - A^{-1} B \frac{\rho}{\lambda} \quad (14)$$

in the $(\rho/\lambda, \mu/\lambda)$ -plane. So long as $A \neq 0$ in (13), the straight line in (13) transforms into the straight line in (14). This observation is very important because the straight line in (11) corresponds to a straight line in the saturation-proxy plot in the $(\rho/\mu, \lambda/\mu)$ -plane. But this line transforms into a straight line in data-sorting plot in the $(\rho/\lambda, \mu/\lambda)$ -plane. In fact the apparent straight line along which the data align themselves in these plots is just this transformed patchy saturation line.

When $A = 0$ in (13) [which seems to happen rarely if ever in the real data examples, but needs to be considered in general], the resulting transformed line will just be one of constant $\rho/\lambda = B^{-1}$, which is a vertical line on the $(\rho/\lambda, \mu/\lambda)$ -plane. The more interesting special case is when $B = 0$, in which situation $\lambda/\mu = A$ or $\mu/\lambda = A^{-1}$. But this case includes that of Gassmann-Domenico for homogeneous mixing of the fluids at low to moderate saturation values. For $B = 0$, on both planes we have horizontal straight lines, but their lengths can differ significantly on the two displays.

Interpreting the data point locations

Data points inside the triangle

The triangle described in Section 3.3 provides rigorous bounds on mechanical properties of porous media. For plots in the $(\rho/\mu, \lambda/\mu)$ -plane such as those included in Figure 1(d) and Figures 3(b), 3(d), and 3(f), some data points lie between the ideal patchy saturation line and the Gassmann ideal lower bound. The relative position of the data points may contain information about the fluid distribution. Consider the case of a core sample that is nearly saturated, above 90% for example. If the weight of the core is used to determine the saturation but the core contains a few gas bubbles, the background saturation will be underestimated and the bubbles themselves represent patches. This is an example of a material having a few isolated patches contained in an otherwise homogeneous partially-saturated background. Such data would plot above but close to the Gassmann curve. In an analogous case for field seismic data, the background saturation may be known from measurements made at lower frequencies or in a nearby region, and it may be possible to use such information to determine the relative volume of patches. For data lying in the middle (*i.e.*, between the bounding curves), some assumptions about fluid distribution could be made and then various estimates about patchy volumes could be applied to different models such as the Hashin-Shtrikman bounds (Hashin and Shtrikman, 1962) or effective medium theories. Exploration of these issues will be the subject of future work.

Data points outside the triangle

The sides of the triangle described above set rigorous boundaries for effects associated with homogeneous saturation and patchy saturation at low frequencies or for situations in which frequency-dependent dispersion can be neglected. However, when the data do not in fact satisfy these assumptions of the theory, plotting the data this way provides an opportunity to

observe and interpret deviations from the behavior predicted by the theory. For example, data which plot above the patchy saturation line represent excessively stiff rock. One possible cause of systematically high stiffness values is frequency-dependent dispersion (Biot, 1956a,b; Biot, 1962; O'Connell and Budiansky, 1977; Mavko and Nur, 1978; Berryman, 1981; McCann and McCann, 1985; Johnson *et al.*, 1987; Norris, 1993; Best and McCann, 1995). Chemical effects, not taken into account in the analysis, might also cause measurements to deviate systematically from predicted behavior. For example, adhesive effects associated with chemical reactions between pore fluid and solid constituents might cause systematically high values. Another consequence of rock-water interactions is softening of intragranular cements. In this case, data for susceptible rocks would systematically plot below the Gassmann line at low saturations. Direct indications from elastic data of rock-water interactions [*e.g.*, see Bonner *et al.* (1997)] may lead to new methods of determining other rock properties controlled by chemical effects, such as the tensile strength.

CONCLUSIONS

We have shown that seismic/sonic velocity data can be transformed to polar coordinates that have quasi-orthogonal dependence on saturation and porosity. This observation is based on the Gassmann-Domenico relations, which are known to be valid at low frequencies. The transformation loses its effectiveness at high frequencies whenever dispersion becomes significant, because then Biot theory and/or other effects play important roles in determining the velocities. So, the simple relations between v_p , v_s , and λ , μ , ρ , and S break down at high frequencies. Our results are, nevertheless, quite encouraging because the predicted relationships seem to work in many cases up to frequencies of 1 kHz, and in a few special cases to still higher frequencies. These results present a straightforward method for obtaining porosity, saturation, and some information about spatial distribution of fluid (*i.e.*, patchy versus homogeneous) in porous rocks and sediments, from compressional and shear wave velocity data alone. These results have potential applications in various areas of interest, including petroleum exploration and reservoir characterization, geothermal resource evaluation, environmental restoration monitoring, and geotechnical site characterization. The methods may also provide physical insight suggesting new approaches to AVO data analysis.

ACKNOWLEDGMENTS

We thank Bill Murphy and Rosemarie Knight for providing access to their unpublished data files. We thank Carmen Mora for helpful comments on the text. We thank Norman H. Sleep for his insight clarifying the significance of our sorting method for plotting seismic data.

REFERENCES

- Aki, K., and Richards, P. G., 1980, *Quantitative Seismology: Theory and Methods*, Vols. I & II, W. H. Freeman and Company, New York.

- Berge, P. A., Bonner, B. P., and Berryman, J. G., 1995, Ultrasonic velocity-porosity relationships for sandstone analogs made from fused glass beads: *Geophysics* **60**, 108–119.
- Berryman, J. G., 1981, Elastic wave propagation in fluid-saturated porous media: *J. Acoust. Soc. Am.* **69**, 416–424.
- Berryman, J. G., 1999, Origin of Gassmann's equations: *Geophysics* **64**, 1627–1629.
- Berryman, J. G., Grechka, V. Y., and Berge, P. A., 1999, Analysis of Thomsen parameters for finely layered VTI media: *Geophys. Prospect.* **47**, 959–978.
- Berryman, J. G., Thigpen, L., and Chin, R. C. Y., 1988, Bulk elastic wave propagation in partially saturated porous solids: *J. Acoust. Soc. Am.* **84**, 360–373.
- Best, A. I., and McCann, C., 1995, Seismic attenuation and pore-fluid viscosity in clay-rich reservoir sandstones: *Geophysics* **60**, 1386–1397.
- Biot, M. A., 1956a, Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range: *J. Acoust. Soc. Am.* **28**, 168–178.
- Biot, M. A., 1956b, Theory of propagation of elastic waves in a fluid-saturated porous solid. II. Higher frequency range: *J. Acoust. Soc. Am.* **28**, 179–1791.
- Biot, M. A., 1962, Mechanics of deformation and acoustic propagation in porous media: *J. Appl. Phys.* **33**, 1482–1498.
- Biot, M. A., and Willis, D. G., 1957, The elastic coefficients of the theory of consolidation: *J. Appl. Mech.* **24**, 594–601.
- Bonner, B. P., Hart, D. J., Berge, P. A., and Aracne, C. M., 1997, Influence of chemistry on physical properties: Ultrasonic velocities in mixtures of sand and swelling clay (abstract): LLNL report UCRL-JC-128306abs, *Eos, Transactions of the American Geophysical Union* **78**, Fall Meeting Supplement, F679.
- Bourbié, T., Coussy, O., and Zinsner, B., 1987, *Acoustics of Porous Media*, Gulf Publishing, Houston, Texas, pp. 56.
- Brown, R. J. S., and Korringa, J., 1975, On the dependence of the elastic properties of a porous rock on the compressibility of the pore fluid: *Geophysics* **40**, 608–616.
- Cadoret, T., 1993, *Effet de la saturation eau/gaz sur les propriétés acoustiques des roches, Étude aux fréquences sonores et ultrasonores*. Ph.D. Dissertation, Université de Paris VII, Paris, France.
- Cadoret, T., Marion, D., and Zinsner, B., 1995, Influence of frequency and fluid distribution on elastic wave velocities in partially saturated limestones: *J. Geophys. Res.* **100**, 9789–9803.
- Cadoret, T., Mavko, G., and Zinsner, B., 1998, Fluid distribution effect on sonic attenuation in partially saturated limestones: *Geophysics* **63**, 154–160.

- Castagna, J. P., and Backus, M. M., 1993, *Offset-Dependent Reflectivity – Theory and Practice of AVO Analysis*, Society of Exploration Geophysicists, Tulsa, OK.
- Castagna, J. P., Batzle, M. L., and Eastwood, R. L., 1985, Relationship between compressional-wave and shear-wave velocities in clastic silicate rocks: *Geophysics* **50**, 571–581.
- Domenico, S. N., 1974, Effect of water saturation on seismic reflectivity of sand reservoirs encased in shale: *Geophysics* **39**, 759–769.
- Dvorkin, J., and Nur, A., 1998, Acoustic signatures of patchy saturation: *Int. J. Solids Struct.* **35**, 4803–4810.
- Endres, A. L., and Knight, R., 1989, The effect of microscopic fluid distribution on elastic wave velocities: *Log Anal.* **30**, 437–444.
- Foster, D. J., Keys, R. G., and Schmitt, D. P., 1997, Detecting subsurface hydrocarbons with elastic wavefields: *Inverse Problems in Wave Propagation*, edited by G. Chavent, G. Papanicolaou, P. Sacks, and W. Symes, Springer, New York, pp. 195–218.
- Gassmann, F., 1951, Über die elastizität poröser medien: *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, **96**, 1–23.
- Harris, J. M., Nolen-Hoeksema, R. C., Langan, R. T., Van Schaack, M., Lazaratos, S. K., and Rector, J. W., III, 1995, High-resolution crosswell imaging of a west Texas carbonate reservoir: Part 1–Project summary and interpretation: *Geophysics* **60**, 667–681.
- Hashin, Z., and Shtrikman, S., 1962, A variational approach to the theory of elastic behaviour of polycrystals: *J. Mech. Phys. Solids* **10**, 343–352.
- Hill, R., 1952, The elastic behaviour of crystalline aggregate: *Proc. Phys. Soc. London A* **65**, 349–354.
- Johnson, D. L., J. Koplik, and R. Dashen, Theory of dynamic permeability and tortuosity in fluid-saturated porous-media: *J. Fluid Mech.* **176**, 379–402 (1987).
- Knight, R., and Nolen-Hoeksema, R., 1990, A laboratory study of the dependence of elastic wave velocities on pore scale fluid distribution: *Geophys. Res. Lett.* **17**, 1529–1532.
- Mavko, G., and Nolen-Hoeksema, R., 1994, Estimating seismic velocities at ultrasonic frequencies in partially saturated rocks: *Geophysics* **59**, 252–258.
- Mavko, G. M., and Nur, A., 1978, The effect of nonelliptical cracks on the compressibility of rocks: *Geophysics* **83**, 4459–4468.
- McCann, C., and McCann, D. M., 1985, A theory of compressional wave attenuation in non-cohesive sediments: *Geophysics* **50**, 1311–1317.
- Murphy, William F., III, 1984, Acoustic measures of partial gas saturation in tight sandstones: *J. Geophys. Res.* **89**, 11549–11559.

- Norris, A. N., 1993, Low-frequency dispersion and attenuation in partially saturated rocks: *J. Acoust. Soc. Am.* **94**, 359–370.
- Nur, A., and Simmons, G., 1969, The effect of saturation on velocity in low porosity rocks: *Earth and Planet. Sci. Lett.* **7**, 183–193.
- O’Connell, R. J., and Budiansky, B., 1977, Viscoelastic properties of fluid-saturated cracked solids: *J. Geophys. Res.* **82**, 5719–5736.
- Ostrander, W. J., 1984, Plane-wave reflection coefficients for gas sands at non-normal angles of incidence: *Geophysics* **49**, 1637–1648.

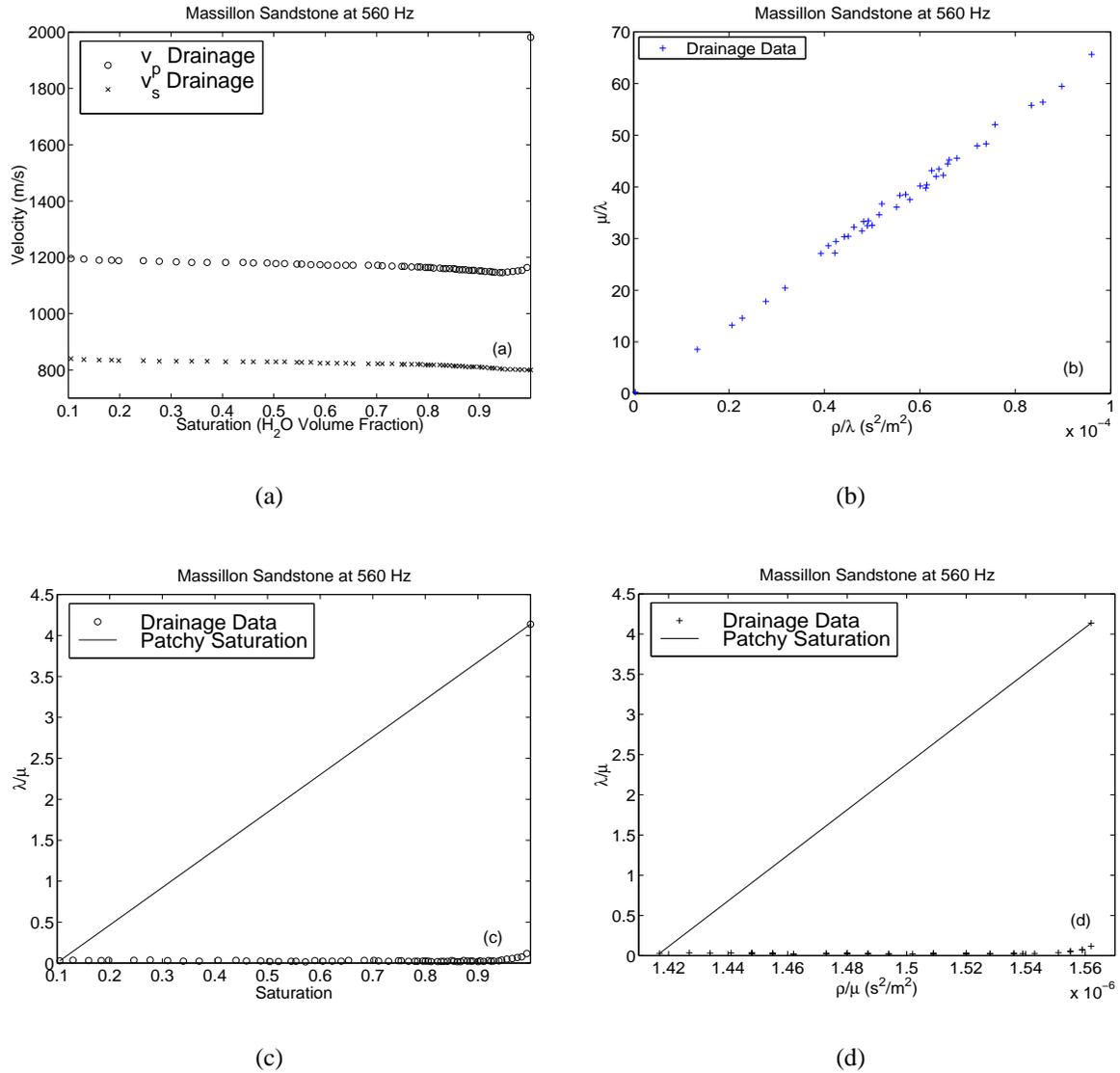


Figure 1: Various methods of plotting 560 Hz Massillon sandstone data of Murphy (1984): (a) Compressional and shear wave velocities as a function of saturation, (b) transform to $(\rho/\lambda, \mu/\lambda)$ -plane, (c) λ/μ versus saturation, (d) transform to $(\lambda/\mu, \rho/\mu)$ -plane. All of these behaviors are anticipated by the Gassmann-Domenico relations for homogeneously mixed fluid in the pores.

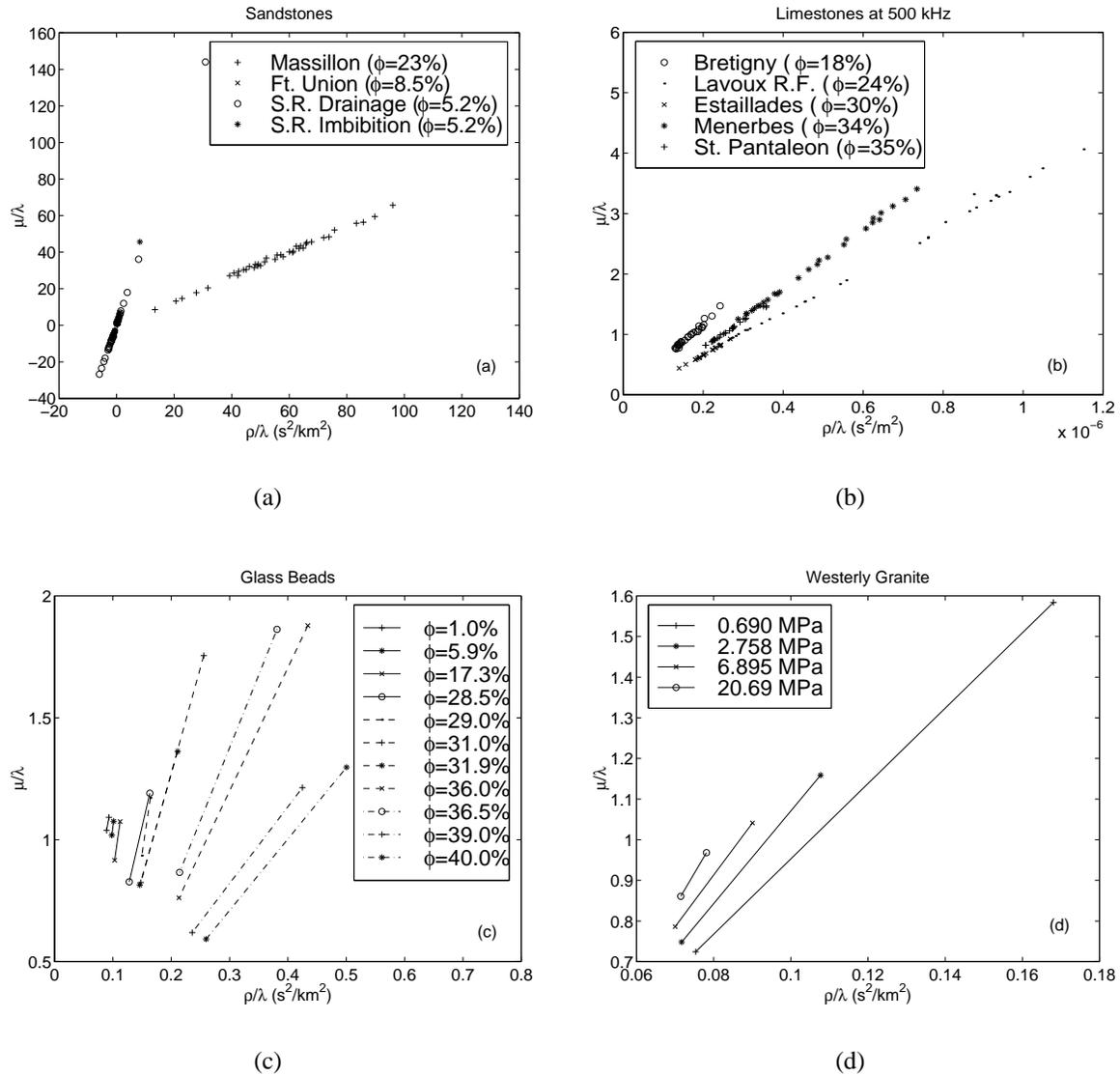


Figure 2: Examples of the correlation of slopes with porosity in the data-sorting plots: (a) three Spirit River (S.R.) sandstone (Knight and Nolen-Hoeksema, 1990) and Massillon and Ft. Union sandstones (Murphy, 1984), (b) five limestones (Cadoret *et al.*, 1998), (c) 11 fused glass-bead samples (Berge *et al.*, 1995), (d) Westerly granite (Nur and Simmons, 1969) at four pressures. The observed trend is that high porosity samples generally have lower slopes than lower porosities on these plots, although there are a few exceptions as discussed in the text. These trends are easily understood since the slopes are determined approximately by the average value of v_s^2 for each material, which is a decreasing function of porosity ϕ .

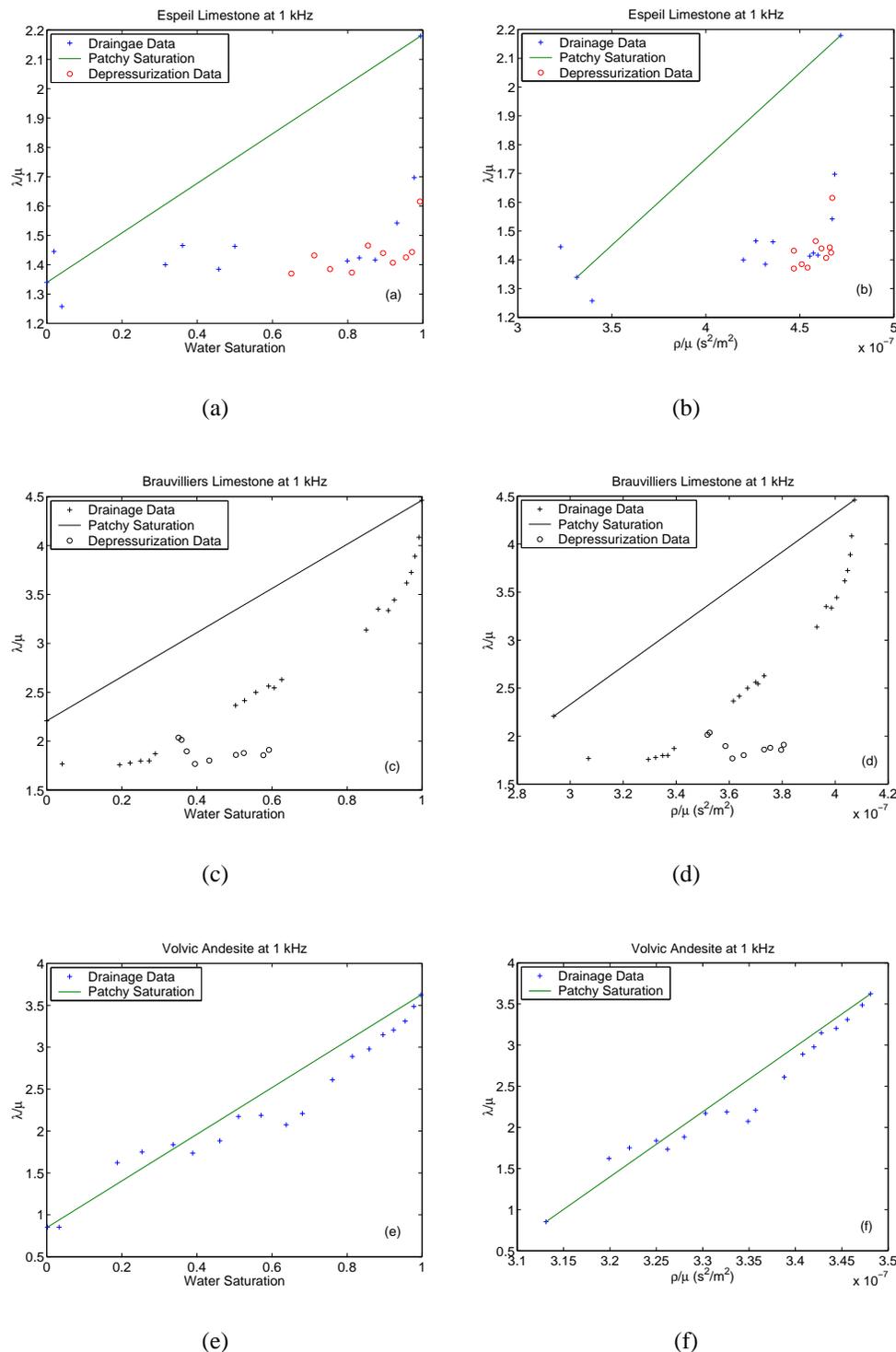


Figure 3: Lamé parameter ratio λ/μ plotted versus (a) saturation and (b) ρ/μ for Espelil limestone, (c) saturation and (d) ρ/μ for Brauvilliers limestone, and (e) saturation and (f) ρ/μ for Volvic andesite. All extensional and shear wave measurements (Cadoret, 1993; Cadoret *et al.*, 1995; 1998) were made at 1 kHz. Note that (a) and (b) indicate homogeneous mixing of liquid and gas, (e) and (f) indicate extremely patchy mixing, while (c) and (d) show an intermediate state of mixing for the drainage data, but more homogeneous mixing for the depressurization data. The plots on the right are saturation-proxy plots, having essentially the same behavior as the plots on the left but requiring only velocity data.

Estimation of AVO attributes sensitivity to velocity uncertainty using forward modeling: a progress report

*Carmen Mora and Biondo Biondi*¹

ABSTRACT

We investigate the sensitivity of AVO attributes to uncertainty in migration velocity in a synthetic dataset. The synthetic data was built using a earth model with typical rock properties from a real North Sea turbidite field. The model includes a thick overburden layer with complex velocity anomalies. We examine the sensitivity of AVO response due to the presence of this complex layer and quantify the influence of migration velocity errors in the AVO signature. Results show that AVO gradient attribute is more sensitive to velocity errors than AVO intercept attribute. For velocity errors up to 5% we see a maximum of AVO intercept errors of 34%, whereas for velocity errors of only 1%, the inversion of AVO gradient attribute has an error of 185%. Further work is needed to evaluate the influence of observed boundary artifacts on these results.

INTRODUCTION

The variation of seismic reflection coefficients with offset can be used as a direct hydrocarbon indicator (Ostrander, 1984; Swan, 1993), which is supported in the AVO analysis theory. AVO analysis requires previous prestack migration of the data, and velocity estimation is a key factor for this imaging problem. Velocity estimation affects the AVO response because it modifies the position of the events and the resulting amplitude values (Grubb and Tura, 1997). Because of the difficulty of estimating velocity models in complex areas, it is important to understand the sensitivity of AVO attributes to variation in velocity models. Mora and Biondi (1999) explore the relationship between velocity uncertainty and AVO-related seismic attributes using a real dataset. A conclusion from that work is that is important to investigate this problem using a synthetic model that allows more control over the data, which is needed to obtain a quantitative measure of the uncertainties.

In this work we do seismic modeling using typical rock properties from a real North Sea turbidite field. As is mentioned in (Avseth et al., 1999), this field has been problematic because of complex sand distribution and non-reservoir sand anomalies. Two of the three most recent exploration wells failed to encounter reservoir sands in locations where poststack seismic amplitudes indicated reservoir sands. Avseth et al. (1999) suggest that AVO analysis in this field can help to discriminate sands from other lithofacies. However, because of the presence of

¹email: cmora@sep.stanford.edu, biondo@sep.stanford.edu

complex velocity anomalies in the overburden, it is desirable to have a estimation of the uncertainty in the AVO response; in other words, how reliable is the lithology discrimination from AVO analysis given the presence of a complex overburden zone?

In this paper, we do forward modeling, simulating an earth model with an overburden that includes complex velocity anomalies. We generate several migration-velocity realizations by introducing coherent percentage velocity errors in the overburden zone of the original velocity model. We migrate the synthetic data using each velocity realization, and measure the variability in the resulting gradient and intercept AVO attributes that results from the velocity error.

ELASTIC MODELING

To investigate the effect of velocity anomalies in AVO attributes, we generated two synthetic datasets using a finite-difference elastic modeling program. Below is a description of the earth models simulated and the resulting synthetic data.

Models

The two 2-D synthetic datasets were computed assuming an earth model that includes

- A 1.8 km thick overburden.
- A 0.2 km cap rock layer (shale).
- A 0.2 km target zone with three different lithologies for comparison (cemented brine sands, cemented oil sands, and tuff).

Figure 1 shows the P-wave velocity for both models. In model 1, the overburden contains two flat layers with constant elastic properties on each layer, whereas model 2 includes a zone of complex velocity anomalies in the overburden. The rock properties for the model were taken from real well logs of a North Sea field. Typical values for different lithologies at this field are listed in Table 1.

Lithology	V_p (km/s)	V_s (km/s)	ρ (g/cm ³)
Shale	2.4	0.95	2.25
Cemented brine sands	3.1	1.55	2.15
Uncemented brine sands	2.6	1.3	2.1
Cemented oil sands	2.9	1.6	2.05
Uncemented oil sands	2.35	1.33	2
Volcanic ash (tuff)	2.75	1.23	2.2
Limestone	4	2	2.4

Table 1. Typical rock properties for different lithologies at the North Sea.

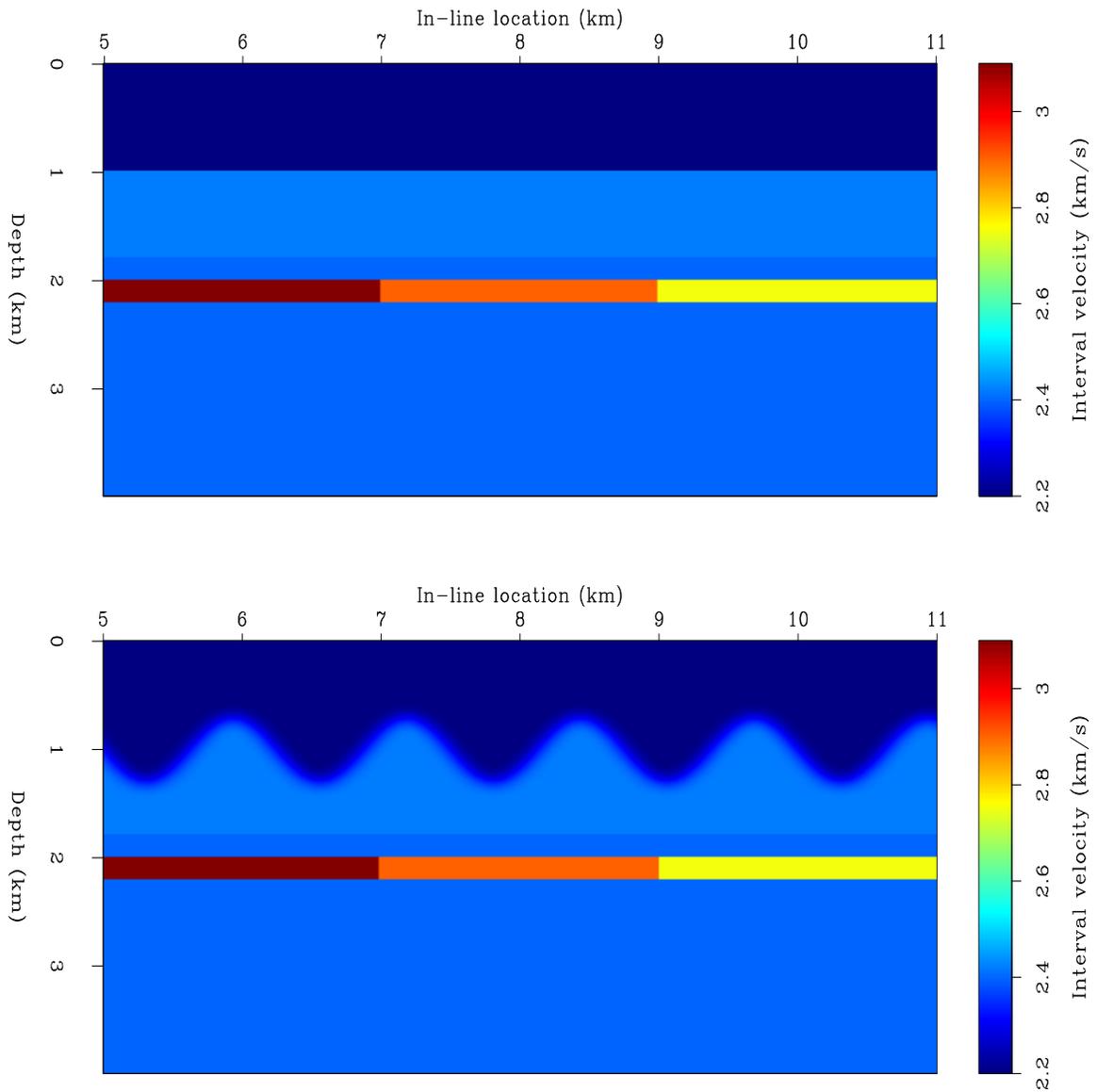


Figure 1: P-wave velocity models used to generate the synthetic data. Model 1 (top): overburden with flat layers, model 2 (bottom): overburden with velocity anomalies. `cmoral-model` [ER]

Average values for overburden properties in the field are $V_p = 2.2$ km/s, $V_s = 0.75$ km/s, $\rho = 2.15$ (g/cm³).

In model 1, overburden properties above the flat interface were taken to be the average values indicated above; overburden properties below the interface were the average values with a 10% increase. For model 2, we introduced lateral velocity anomalies by including a smoothed sinusoidal interface between the two layers.

Synthetic seismograms

The synthetic data was generated using an explosive source and a Ricker2 wavelet with a fundamental frequency of 22.5 Hz. The source/receiver offsets ranged from 16 m (minimum offset) to 3.6 km (maximum offset). Figure 2 shows a shot gather at in-line location=5 km for each model; note that the events in model 2 are not perfect hyperboles due to the lateral velocity variations.

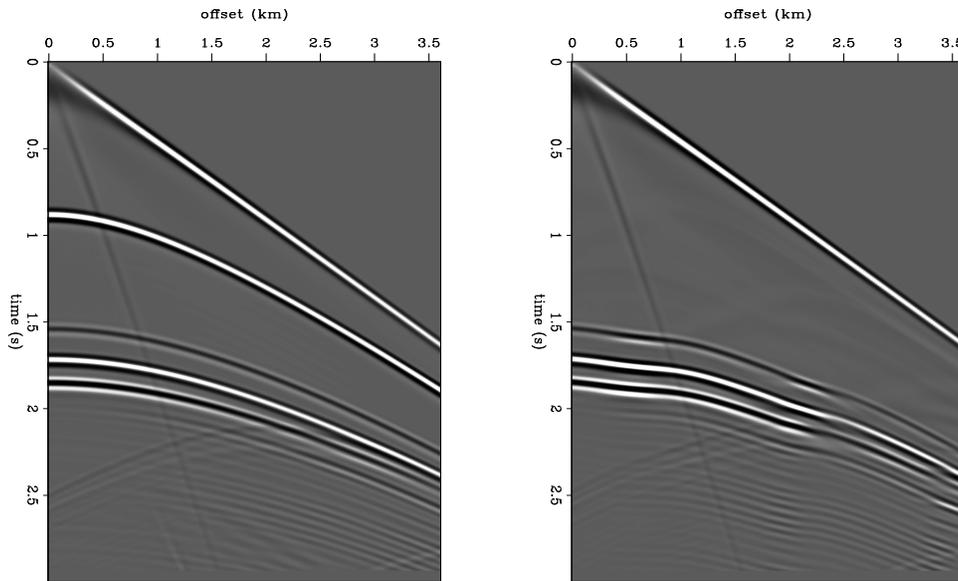


Figure 2: Shot gather at in-line location=5 km. Left: model 1, Right: model 2 cmora1-shot
[CR]

Preprocessing

Divergence correction, coherent noise suppression, and CMP sorting were applied to the data before 2-D prestack depth migration. In order to compensate for spherical divergence, we scaled the data in the time axis using a function $\text{trace}(t) = \text{trace}(t) * t$. Coherent noise suppression was also applied to eliminate the P-wave and S-wave first arrivals. To eliminate the P-wave first arrival, we applied a linear outer mute to the data. The S-wave arrival was suppressed by applying a 2-D dip filter. Finally, the data was windowed to extract the CMPs with maximum

offset coverage and sorted by CMP location. Figure 3 shows the zero-offset section of the resulting data.

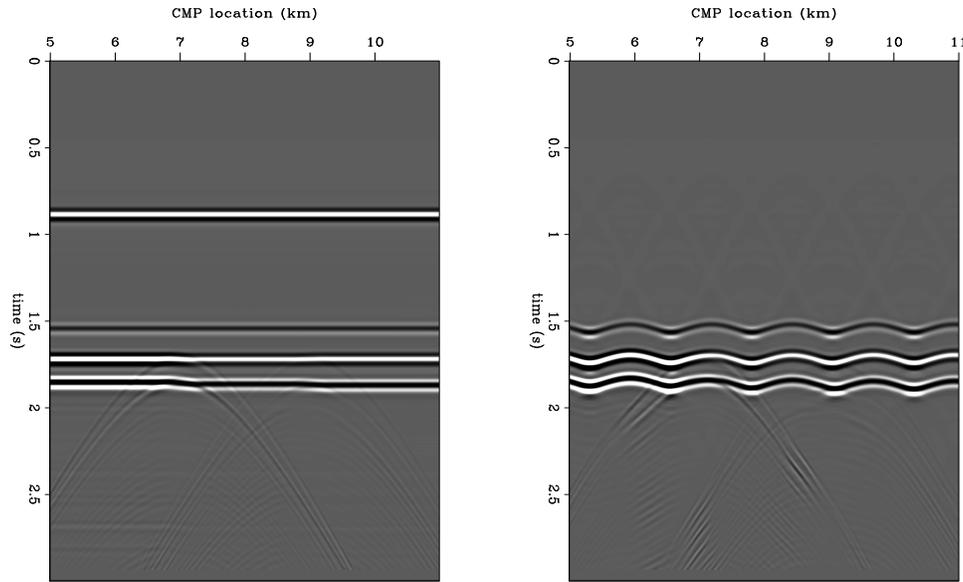


Figure 3: Zero-offset section of the resulting synthetic data. Left: model 1, Right: model 2
cmora1-zero-off [CR]

PRESTACK MIGRATION

From the original velocity model, we generated several velocity model realizations by applying a percentile perturbation at the overburden zone. Using each velocity realization, we applied a 2-D prestack wave-equation migration (Prucha et al., 1999) to the synthetic data. The resulting image is a function of the offset ray parameter p_{hx} , which is related to the aperture angle θ , the dip ϕ along the in-line direction, and the velocity function $V(z, m)$, as follows:

$$p_{hx} = \frac{2 \sin \theta \cos \phi}{V(z, m)} \quad (1)$$

Figure 4 shows the result of applying prestack wave-equation migration to the synthetic data, using the original velocity models (0% perturbation).

AVO INVERSION

The physical relation between the variation of reflection/transmission coefficients with incident angle (and offset) and rock parameters has been widely investigated. This relation is established in the Zoeppritz equations, which relate reflection and transmission coefficients for plane waves and elastic properties of the medium. Because of the nonlinearity of the Zoeppritz equations, several approximations have been generated, such as those presented by

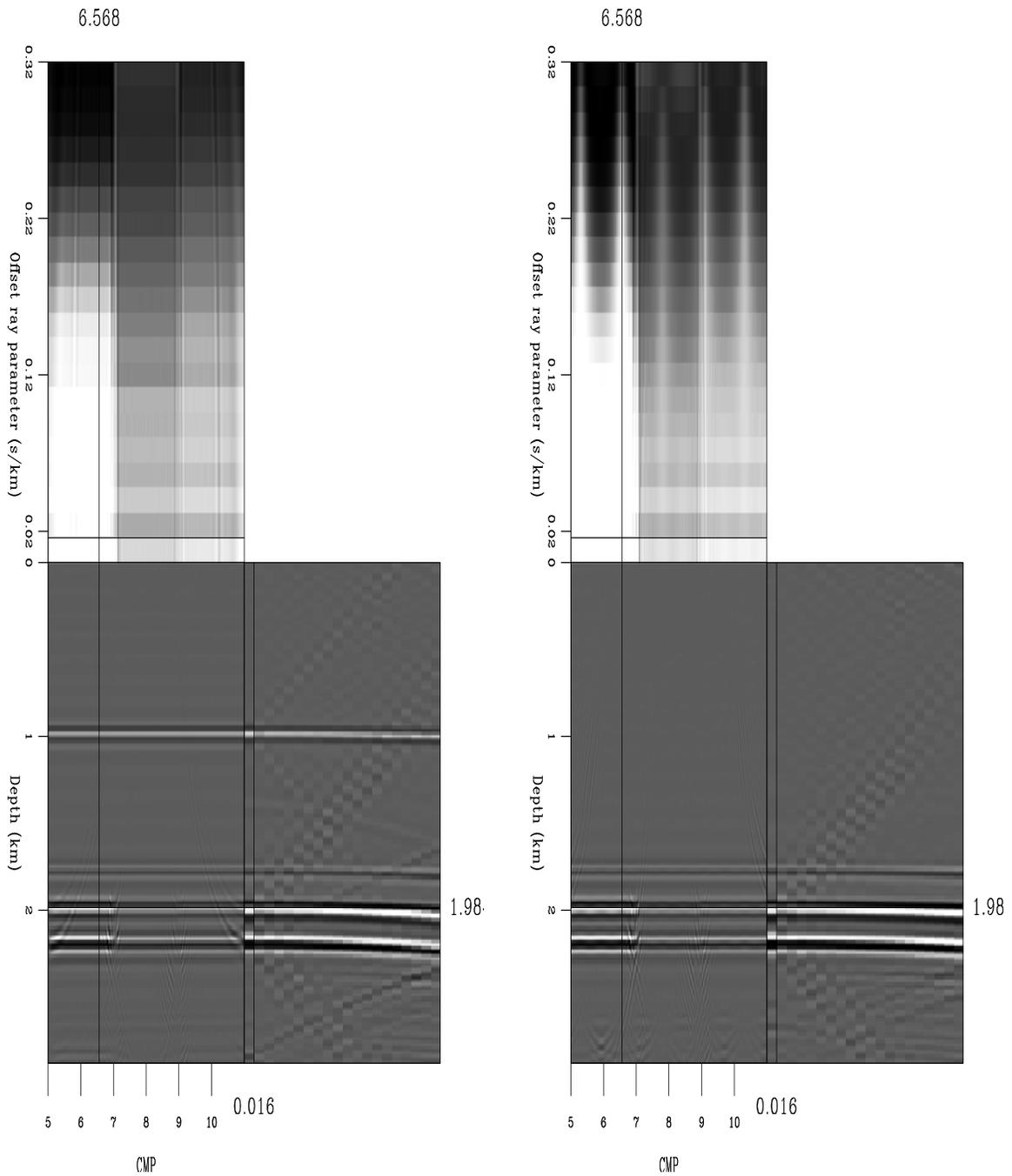


Figure 4: Result of applying prestack wave-equation migration to the synthetic data using the original velocity model. Left: model 1, right: model 2. `cmora1-mig` [CR]

Aki and Richards (1997) and Shuey (1985). The simplified versions of Zoeppritz equations allow the computation of AVO inversion to estimate elastic parameters from the observed reflection amplitude variation with angle. Equation (2) from Castagna and Smith (1994) is a version of Shuey's approximation for the P-wave reflection coefficient as a function of angle of incidence, which is linear in $\sin^2 \theta$. This equation characterizes the reflection coefficient, at normal incidence, and at intermediate angles ($0 < \theta < 30$ degrees),

$$R(\theta) \approx A + B \sin^2 \theta \quad (2)$$

$$A = \left(\frac{\Delta V_p}{V_p} + \frac{\Delta \rho}{\rho} \right) / 2 \quad (3)$$

$$B = \left(-2 \frac{V_s^2}{V_p^2} \frac{\Delta \rho}{\rho} + \frac{\Delta V_p}{2V_p} - 4 \frac{V_s^2}{V_p^2} \frac{\Delta V_s}{V_s} \right) \quad (4)$$

where

$$V_p = (V_{p2} + V_{p1})/2,$$

$$V_s = (V_{s2} + V_{s1})/2,$$

$$\rho = (\rho_2 + \rho_1)/2,$$

$$\Delta V_p = V_{p2} - V_{p1},$$

$$\Delta V_s = V_{s2} - V_{s1},$$

$$\Delta \rho = \rho_2 - \rho_1.$$

The normal incident term, A, is commonly referred to as the AVO intercept attribute, the intermediate angles term, B, is referred to as the AVO gradient attribute. We use this approximation to invert for the intercept and gradient AVO attributes from the observed reflection amplitude variation with angle in the angle-domain common image gathers (CIG). In this domain, we pick the amplitude values at the reflector of interest and fit the amplitude versus $\sin^2 \theta$ to a best straight-line approximation using a least-squares curve fitting method.

Providing a reference for the expected AVO response for the shale/brine, shale/oil, and shale/tuff interface, Figure 5 shows the P-wave reflection coefficient from the exact Zoeppritz equations. At the near offset we expect a similar reflection coefficient (similar intercept attribute) for the shale/oil and the shale/tuff interfaces because of similar acoustic impedance; however this ambiguity can be resolved by the different ratio between V_s and V_p (different gradient attribute). Although this calculation is valid only for a 2-layer model, it will be a reference for the expected tendency in the modeled data. Deviation from this tendency should be due to modeling effect, overburden effect, migration operator effect, velocity anomalies effect, and migration-velocity errors. We examined the modeling effect using a 1-D, 2-layer synthetic model, the overburden and migration operator effects using model 1 (overburden with flat interface), and then we use model 2 (overburden with sinusoidal interface) to understand the effect of velocity anomalies and migration-velocity errors.

Figure 5: P-wave reflection coefficient from Zoeppritz equations
cmora1-zoeppritz [CR]

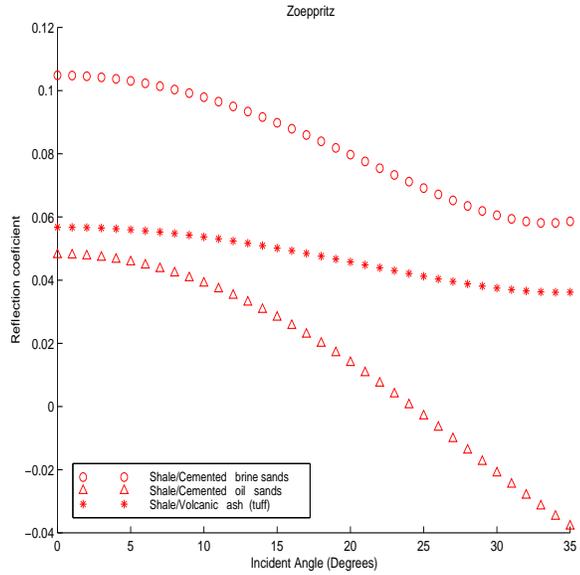
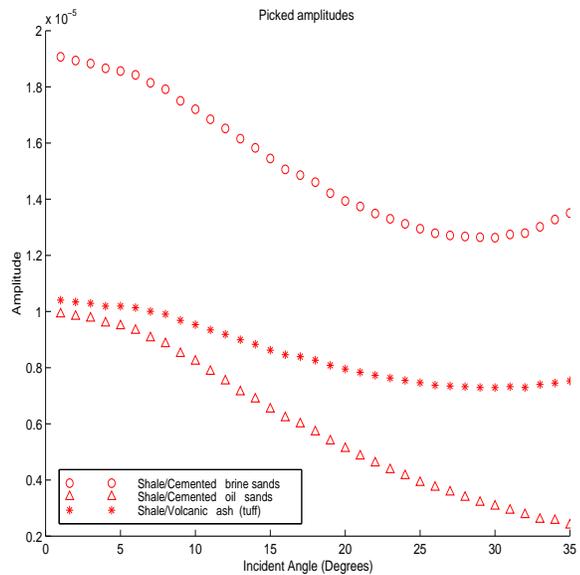


Figure 6: Picked amplitudes for a 2-layer modeling case
cmora1-2laypick [CR]



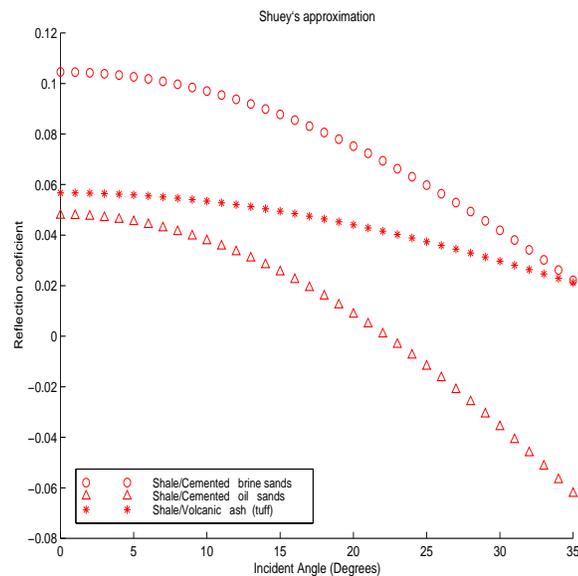
Modeling effect

Illustrating the effect of the modeling in the amplitudes, Figure 6 shows the picked amplitudes (without any calibration) corresponding to a 2-layer modeling case (no overburden). Comparing this figure with Figure 5, we can see a very good qualitative agreement between the relative amplitudes of the different interfaces (shale/brine, shale/oil, and shale/tuff) and the corresponding reflection coefficients. We can see some difference for incident angles > 25 degrees; for this reason, and because Shuey's approximation is valid for incident angles < 30 degrees (see Figure 7), we will restrict the data to be used for the AVO inversion up to 25 degrees of incident angle.

We can also notice that Zoeppritz equation predicts a polarity change for the shale/oil case. We observed the same polarity change in the data (see Figure 8); however, this is not present in the picked amplitudes because we are using an automated picking program that picks the maximum absolute value in a moving window in time. We consider the automatic picking still valid in this modeling case because the polarity change is evident in the data after offset > 2 km, which corresponds to angles > 27 degrees, and we used angles up to 25 degrees for the AVO inversion. We did not use the automatic picking on migrated data, where the events are supposed to be flattened by the migration; rather, in this case, we followed the constant time corresponding to the event.

Figure 7: Shuey's approximation of P-wave reflection coefficient

`cmora1-shuey` [CR]



Overburden effect

We used model 1 to understand the effect of the overburden in the resulting amplitudes. We compared the amplitudes of the 2-layer case to the amplitudes in model 1, which includes an overburden zone with 2 flat layers and a target zone encased in shale (see top of Figure 1). The picked amplitudes at the top of the target zone (2 km depth) are plotted in Figure 9; note the good agreement with the results illustrated in Figure 6 corresponding to the 2-layer case.

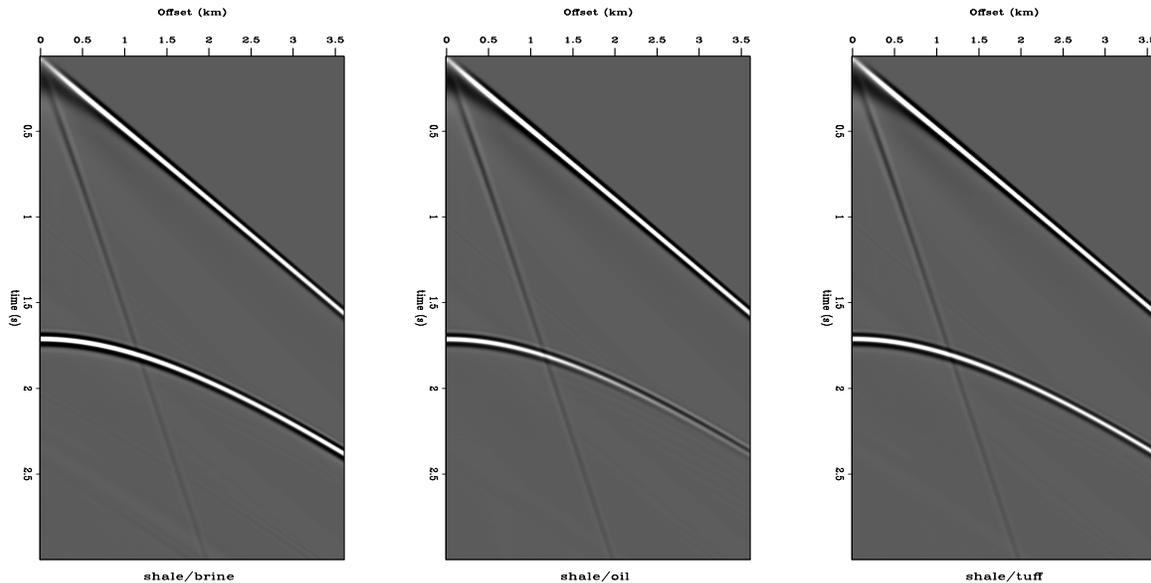


Figure 8: Modeling of shot gather for a shale/brine, shale/oil, and shale/tuff interfaces (1-D, 2-layer model). Notice the change in polarity for offset > 2 km in the shale/oil case `cmora1-2lay_all` [CR]

The amplitudes in model 1 are slightly higher, which can be expected because in this case the reflector appears at a higher time and the spherical divergence correction that we are applied increases the amplitudes more at a higher time (without the velocity taken into account). To confirm this, we picked the amplitudes before any preprocessing and they are identical at the near offset. We also expected some differences at the far offsets due to the AVO effect of the additional layers in model 1 (overburden effect).

Using Equation (2), we calculated the intercept (A) and gradient (B) attributes from the picked amplitudes in model 1 by fitting the amplitude versus $\sin^2\theta$ values to a straight-line approximation using a least-squares curve fitting method. Figure 11 shows the crossplot of the resulting intercept and gradient attributes; note the good quantitative agreement with Figure 10, which illustrates the theoretical values from Shuey's approximation.

Migration effect

We applied the 2-D prestack wave-equation to the synthetic data corresponding to model 1 (overburden with flat interface) using the original velocity model and picked the amplitudes at the top of the target zone in the resulting angle domain CIG. Figure 13 shows the resulting amplitudes as a function of the aperture angle at CIG locations corresponding to each lithology. The aperture angle was calculated from the offset ray parameter p_{hx} using Equation (1), with $\phi = 0$ and $V(z, m) =$ the interval velocity at the interface. We can observe some irregularities in the amplitudes values after migration. Figure 12 shows the intercept and gradient attribute calculated for each CIG. We can notice some artifacts in the data related with boundary effects;

Figure 9: Picked amplitudes from model 1 `cmora1-model1pick` [CR]

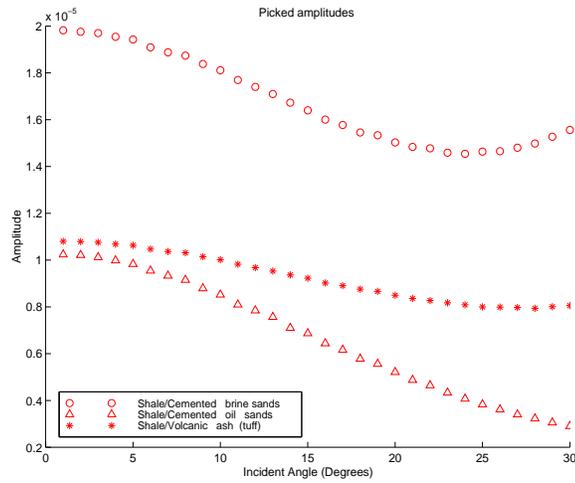


Figure 10: Intercept versus Gradient crossplot from Shuey's approximation of P-wave reflection coefficient `cmora1-ABshuey` [CR]

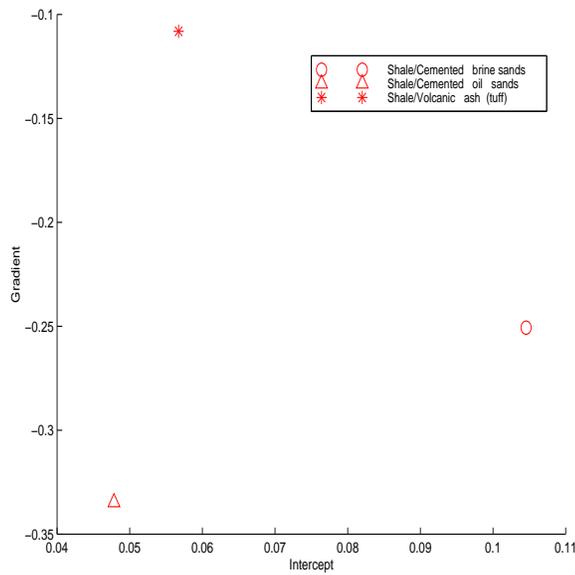
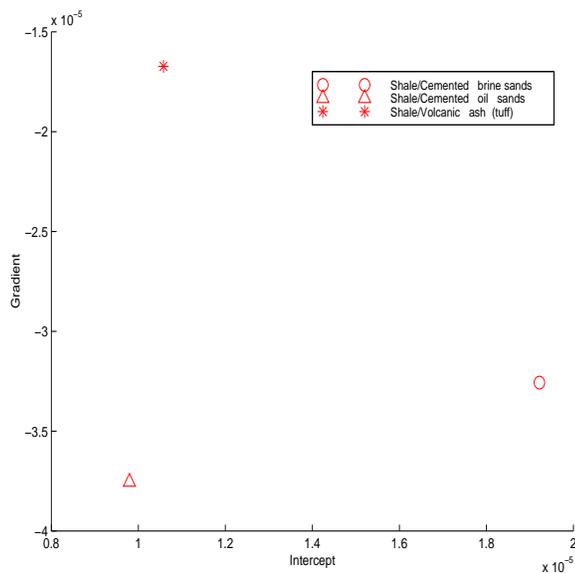


Figure 11: Intercept versus Gradient crossplot from picked amplitudes in data model 1 before migration `cmora1-ABmodel1` [CR]



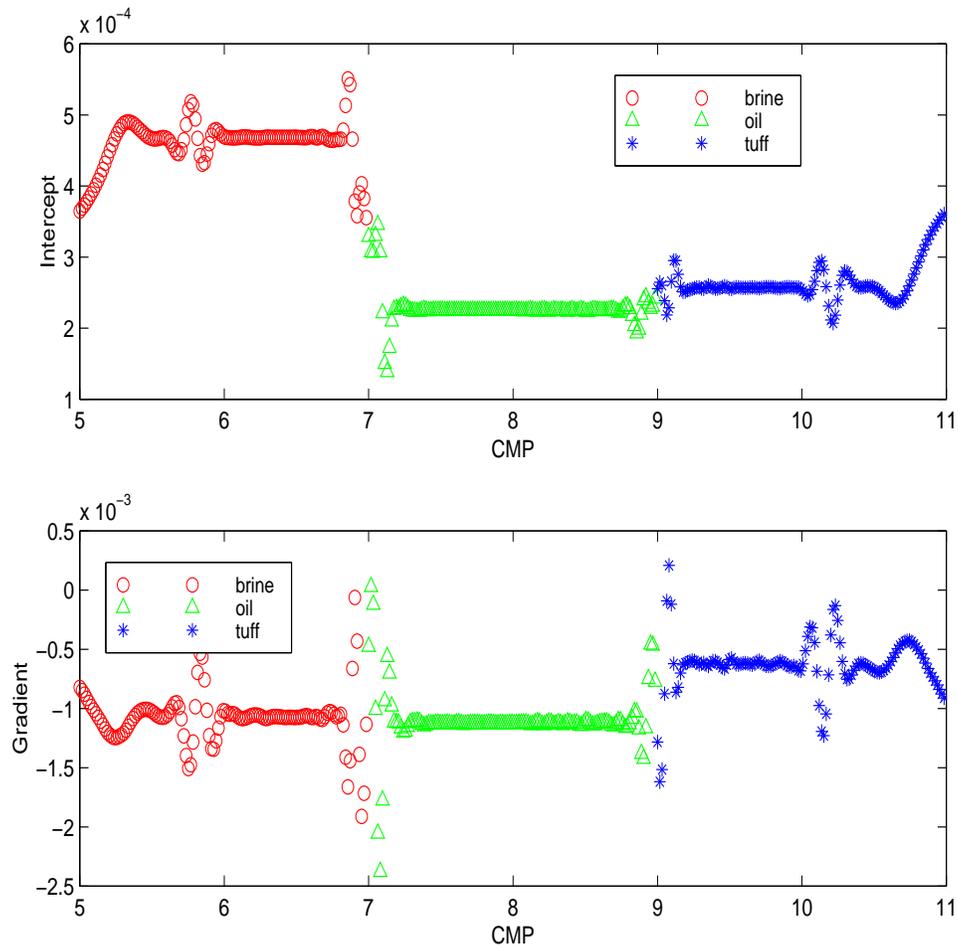


Figure 12: Variability of intercept and gradient attribute along the different CIGs of model 1
cmora1-ABmodel1_var [CR]

these artifacts are very strong at the edges of the model and at the lateral interfaces between the different lithologies at the target zone. Even though the CIGs used for Figure 13 were taken from locations where we noted less variability, these artifacts could be affecting the amplitudes.

Figure 14 shows the crossplot of the inverted intercept and gradient attributes for the same CIG locations used for Figure 13. The relative intercept and attribute values for the different lithologies are in good agreement with the expected tendency, however, we can note from Figure 12 that this will not be the case if the CIG's are chosen close to the boundary artifacts.

Figure 13: Picked amplitudes from CIG of model 1 `cmora1-model1migpick` [CR]

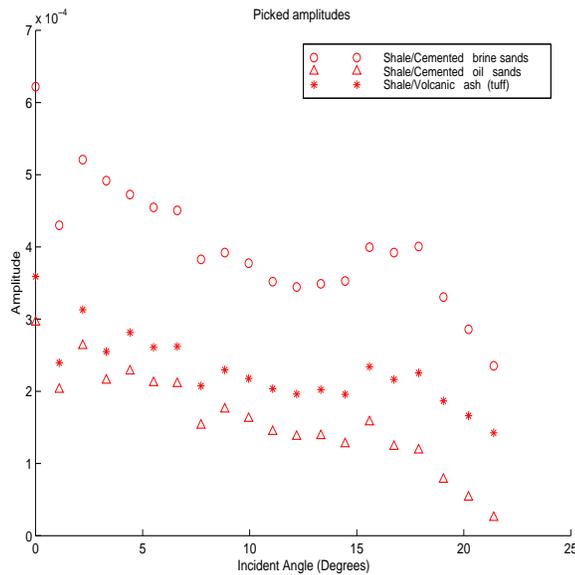
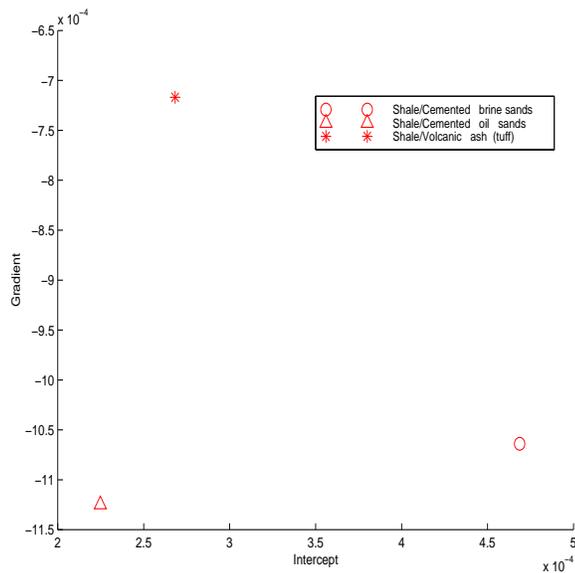


Figure 14: Intercept versus Gradient crossplot from picked amplitudes of model 1 after migration `cmora1-ABmodel1mig` [CR]



Velocity anomalies effect

In this case, we applied the 2-D prestack wave-equation to the synthetic data corresponding to model 2 (overburden with sinusoidal interface) using the original velocity model. We modeled and migrated the data using 1 more km on each side of the inline axis to avoid the edge artifacts at the boundary of the model, but the edge artifacts corresponding to the lateral boundaries between brine-oil and oil-tuff lithologies are still present. Figure 15 shows the picked the amplitudes at the top of the target zone in the migrated CIG. Note how the intercept and gradient change follow the sinusoidal velocity anomalies. Figure 16 shows the picked amplitudes

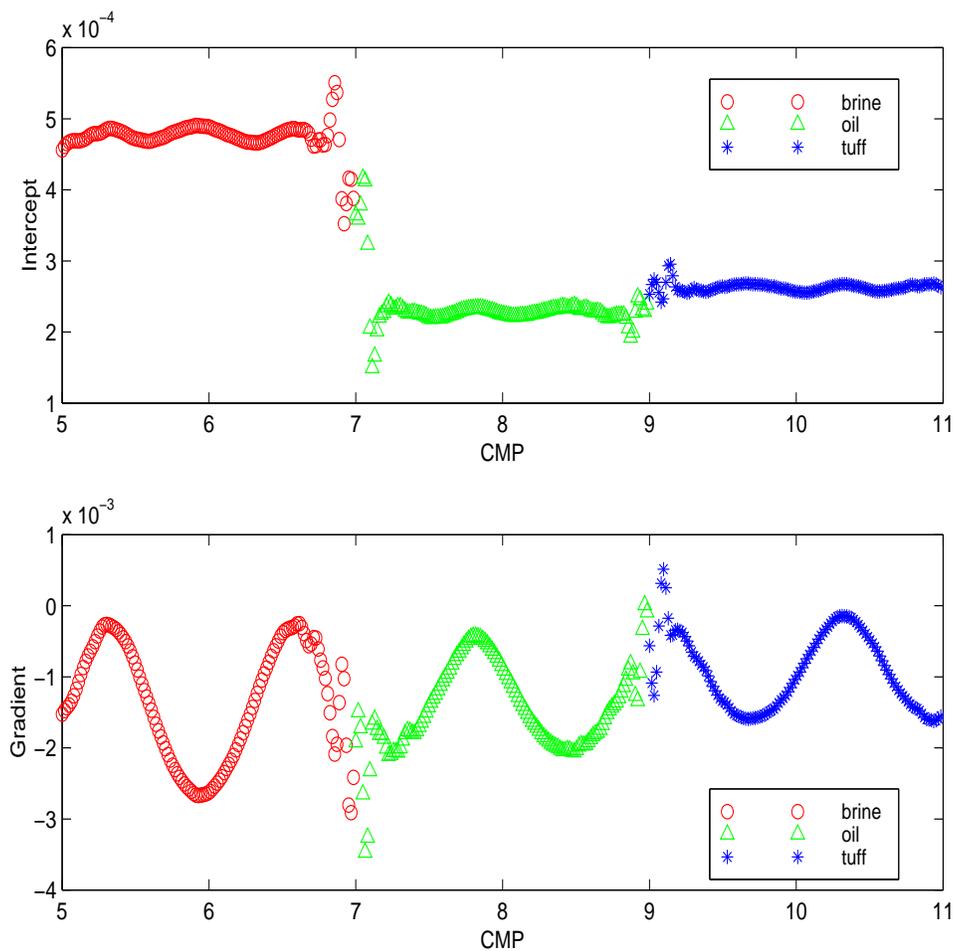
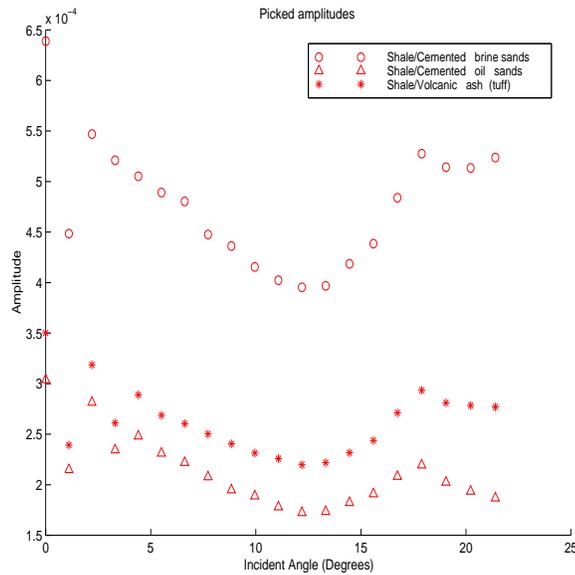


Figure 15: Impact of velocity anomalies in intercept and gradient attribute cmora1-AB_var
[CR]

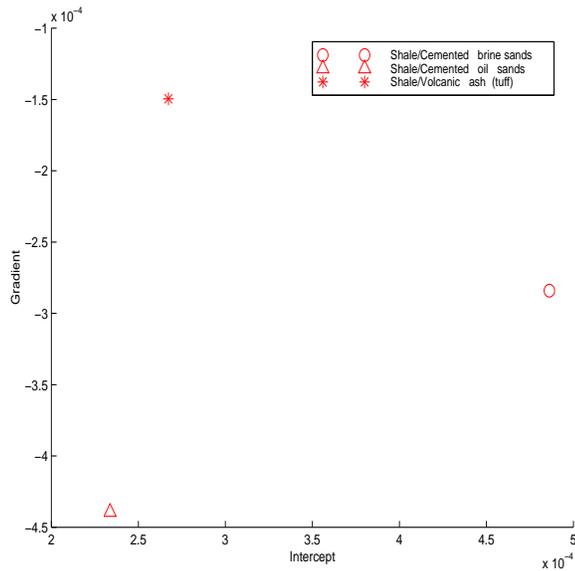
at the CIG locations, which correspond to a valley of the sinusoidal irregularities (where the event was observed to be flatter) for each lithology (positions 6.576, 7.824, and 10.304), and Figure 17 shows the crossplot of the corresponding intercept and gradient attributes. In both cases, we can observe a good agreement with the expected tendencies. We also calculated the intercept and gradient attributes at CIGs locations which correspond to a peak of the sinusoidal irregularities; in this case, the relative gradient value for the shale/brine interface is higher than

Figure 16: Picked amplitudes from CIG of model 2
cmora1-model2migpick [CR]



expected.

Figure 17: Intercept versus Gradient crossplot from picked amplitudes of model 2 after migration
cmora1-ABmodel2mig [CR]



Velocity errors effect

Using the synthetic data corresponding to model 2, we generated several migration-velocity realizations by introducing coherent percentage velocity errors at the overburden zone of the original velocity model. Using each velocity realization, we applied 2-D prestack wave-equation migration to the synthetic data; we applied an additional residual moveout correction and picked the resulting amplitudes. Figure 18 shows the crossplot of the intercept and gradient attributes at CIGs location which correspond to a valley of the sinusoidal irregularities; the

size in the plot symbol increases as the velocity error increases. We can note that the intercept attribute is much less sensitive to velocity errors than the gradient attribute. Figure 19 shows the errors in the inverted attributes as a function of the velocity errors used in the migration. We can see that the maximum AVO intercept error is 34% for velocity errors up to 5% (tuff case), whereas for velocity errors of only 1%, the inversion of AVO gradient attribute (brine case) has an error of 185%.

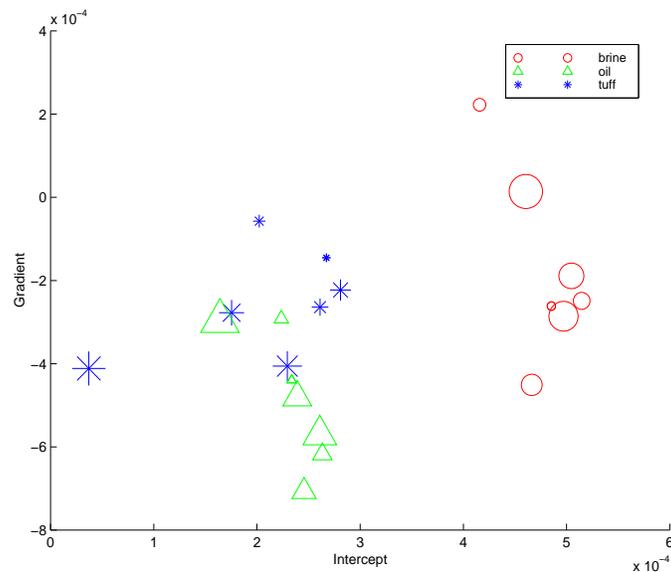


Figure 18: Impact of velocity errors in Intercept versus Gradient crossplot cmora1-mig_vel [CR]

CONCLUSION

We examined the sensitivity of the AVO response due to the presence of a overburden with complex velocity anomalies using a synthetic data set. We observed that AVO attributes calculated after prestack depth migration using the true velocity model are sensitive to the velocity anomalies. Introducing errors in the migration-velocity, we found that the AVO gradient attribute is much more sensitive to velocity errors than AVO intercept attribute. For velocity errors up to 5%, we can see a maximum of AVO intercept errors of 34%, whereas for velocity errors of only 1%, the inversion of AVO gradient attribute has an error of 185%. These results are specific for the synthetic data used; different results could be obtained by modeling different velocity anomalies.

We observed some boundary artifacts in the modeled data and we noted that amplitude values after migration are more sensitive to these boundary artifacts than amplitude values before migration. These boundary artifacts become worse when we introduce velocity errors in the migration-velocity. We need to do further work to evaluate the influence of boundary artifacts on the amplitudes; we also would like to compare the results using other migration methods, such as Kirchhoff prestack migration.

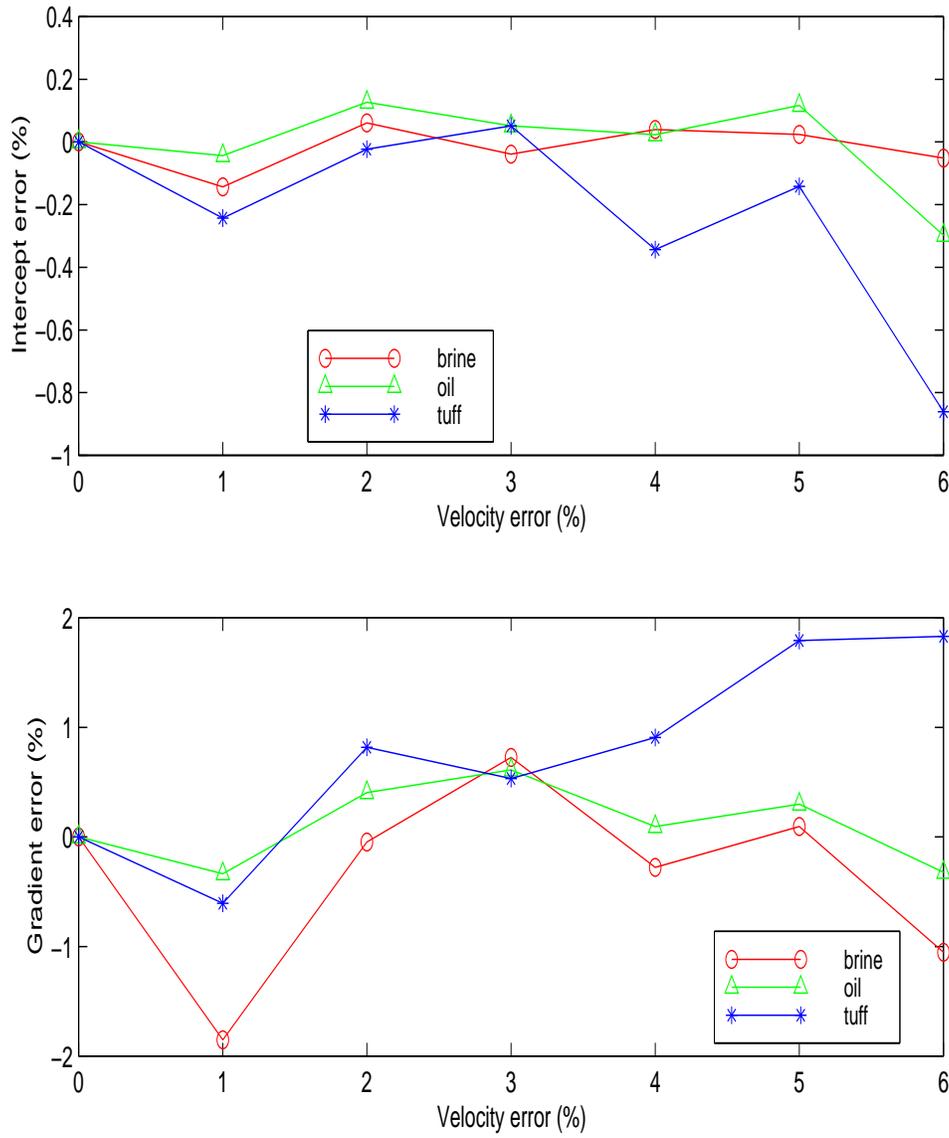


Figure 19: Impact of velocity errors in Intercept and gradient attributes cmora1-AB_error
[CR]

ACKNOWLEDGMENTS

We gratefully acknowledge James Rickett and Antoine Guitton for their many helpful suggestions and discussions. We also would like to thank Per Avseth for providing the earth model properties used for the modeling and Robert Clapp for his support during the development of this work.

REFERENCES

- Aki, K. I., and Richards, P. G., 1997, *Quantitative seismology: Theory and methods*: W. H. Freeman and Company, New York.
- Avseth, P., Mavko, G., Mukerji, T., and Tyssekvam, J., 1999, Integrating seismic lithofacies prediction and depositional geometry analysis for reservoir delineation in a north sea turbidite field: 69th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 752–755.
- Castagna, J. P., and Smith, S. W., 1994, Comparison of avo indicators: A modeling study: *Geophysics*, **59**, no. 12, 1849–1855.
- Grubb, H., and Tura, A., 1997, Interpreting uncertainty measures for AVO migration/inversion: 67th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 210–213.
- Mora, C., and Biondi, B., 1999, Seismic attribute sensitivity to velocity uncertainty: *SEP-102*, 157–176.
- Ostrander, W. J., 1984, Plane-wave reflection coefficients for gas sands at nonnormal angles of incidence: *Geophysics*, **49**, no. 10, 1637–1648.
- Prucha, M. L., Biondi, B. L., and Symes, W. W., 1999, Angle-domain common image gathers by wave-equation migration: *SEP-100*, 101–112.
- Shuey, R. T., 1985, A simplification of the Zoeppritz equations: *Geophysics*, **50**, no. 4, 607–614.
- Swan, H. W., 1993, Properties of direct AVO hydrocarbon indicators *in* Castagna, J. P., and Backus, M. M., Eds., *Offset dependent reflectivity — Theory and practice of AVO analysis*: Society of Exploration Geophysics, 78–92.

SEP ARTICLES PUBLISHED OR IN PRESS

- Bachrach, R., and Rickett, J., 1999, Ultra shallow seismic reflection in depth: examples from 3D ultra shallow survey with application to joint seismic and GPR imaging: 69th Ann. Internat. Meeting, Soc. Expl. Geophys., Expanded Abstracts, 488–491.
- Biondi, B., and Sava, P., 1999, Wave-equation migration velocity analysis: 69th Ann. Internat. Meeting, Soc. Expl. Geophys., Expanded Abstracts, 1723–1726.
- Biondi, B., Fomel, S., and Chemingui, N., 1998, Azimuth moveout for 3-D prestack imaging: *Geophysics*, **63**, no. 2, 574–588.
- Biondi, B., 2000, 3-D wave-equation prestack imaging under salt: 70th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 906–909.
- Brown, M., and Claerbout, J., 2000, A pseudo-unitary implementation of the Radial Trace Transform: 70th Ann. Internat. Meeting, Soc. Expl. Geophys., Expanded Abstracts, –.
- Brown, M., and Clapp, R., 2000, (t,x) domain, pattern-based ground roll removal: 70th Ann. Internat. Meeting, Soc. Expl. Geophys., Expanded Abstracts, submitted.
- Chemingui, N., and Biondi, B., 1999, Data regularization by inversion to common offset (ICO): 69th Ann. Internat. Meeting, Soc. Expl. Geophys., Expanded Abstracts, 1398–1401.
- Claerbout, J., 1998, Multidimensional recursive filters via a helix: *Geophysics*, **63**, 1532–1541.
- Clapp, R. G., and Biondi, B., 2000, Tau domain migration velocity analysis using angle CRP gathers and geologic constraints: 70th Ann. Internat. Meeting, Soc. Expl. Geophys., Expanded Abstracts, submitted.
- Crawley, S., Clapp, R., and Claerbout, J., 1999, Interpolation with smoothly nonstationary prediction-error filters: 69th Ann. Internat. Meeting, Soc. Expl. Geophys., Expanded Abstracts, 1154–1157.
- Guitton, A., and Symes, W. W., 1999, Robust and stable velocity analysis using the Huber function: 69th Ann. Internat. Meeting, Soc. Expl. Geophys., Expanded Abstracts, 1166–1169.
- Prucha, M., Biondi, B., and Symes, W., 1999, Angle-domain common-image gathers by wave-equation migration: 69th Ann. Internat. Meeting, Soc. Expl. Geophys., Expanded Abstracts, 824–827.
- Rickett, J., and Claerbout, J., 1999a, Acoustic daylight imaging via spectral factorization: Helioseismology and reservoir monitoring: 69th Ann. Internat. Meeting, Soc. Expl. Geophys., Expanded Abstracts, 1675–1678.
- Rickett, J., and Claerbout, J., 1999b, Acoustic daylight imaging via spectral factorization: Helioseismology and reservoir monitoring: *The Leading Edge*, **18**, 957–960.

- Rickett, J., and Claerbout, J., 1999c, Time-distance helioseismology, spectral factorization and hydrocarbon reservoir monitoring: Time-distance helioseismology, spectral factorization and hydrocarbon reservoir monitoring: The 9th SOHO Workshop, "Helioseismic Diagnostics of Solar Convection and Activity".
- Rickett, J., 2000, Efficient 3-D wavefield extrapolation with Fourier finite-differences and helical boundary conditions: 62nd Mtg., Eur. Assoc. Geosc. Eng., Abstracts, accepted for presentation.
- Sava, P., and Fomel, S., 2001, 3-D travelttime computation by Huygens wavefront tracing: *Geophysics*, **66**, no. 3, in press.
- Sava, P., 2000, Prestack Stolt residual migration for migration velocity analysis: 70th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, submitted.
- Vaillant, L., Calandra, H., Sava, P., and Biondi, B., 2000a, 3-D wave-equation imaging of a North Sea dataset: Common-azimuth migration + residual migration: 70th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 874–877.
- Vaillant, L., Sava, P., Biondi, B., and Calandra, H., 2000b, 3-D wave-equation migration of a common-azimuth North Sea dataset: 62nd Mtg., Eur. Assoc. Geosc. Eng., Abstracts, 231–235.

SEP PHONE DIRECTORY

Name	Phone	Login Name
Berryman, James	723-1250	berryman
Biondi, Biondo	723-1319	biondo
Brown, Morgan	723-6006	morgan
Claerbout, Jon	723-3717	jon
Clapp, Robert	723-0253	bob
Crawley, Sean	723-0253	sean
Fomel, Sergey	725-1334	sergey
Gratwick, Douglas	723-9282	doug
Guitton, Antoine	723-6007	antoine
Lau, Diane	723-1703	diane
Mora, Carmen	723-5911	cmora
Prucha, Marie	723-0253	marie
Rickett, James	723-6007	james
Rosales, Daniel	723-5647	daniel
Sava, Paul	723-0463	paul
Vaillant, Louis	723-0463	louis

SEP fax number: (650) 723-0683

E-MAIL

Our Internet address is “sep.stanford.edu”; i.e., send Jon electronic mail with the address “jon@sep.stanford.edu”.

WORLD-WIDE WEB SERVER INFORMATION

The current login name for the sponsors’ area on our world-wide web server (<http://sepwww.stanford.edu>) is “sepsponsor2”. The current password is: ss2&sep2

Research Personnel

James G. Berryman received a B.S. degree in physics from Kansas University (Lawrence) in 1969 and a Ph.D. degree in physics from the University of Wisconsin (Madison) in 1975. He subsequently worked on seismic prospecting at Conoco. His later research concentrated on sound waves in rocks – at AT&T Bell Laboratories (1978-81) and in Earth Sciences at Lawrence Livermore National Laboratory (1981-). He is currently a physicist in the Earth and Environmental Sciences Directorate at LLNL. Continuing research interests include seismic and electrical methods of geophysical imaging and waves in porous media containing fluids. He is a member of APS, AGU, SEG, and EEGS.



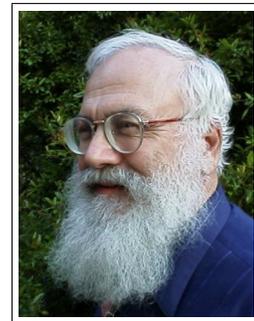
Biondo L. Biondi graduated from Politecnico di Milano in 1984 and received an M.S. (1988) and a Ph.D. (1990) in geophysics from Stanford. SEG Outstanding Paper award 1994. During 1987 he worked as a Research Geophysicist for TOTAL, Compagnie Française des Pétroles in Paris. After his Ph.D. at Stanford Biondo worked for three years with Thinking Machines Co. on the applications of massively parallel computers to seismic processing. After leaving Thinking Machines Biondo started 3DGeo Development, a software and service company devoted to high-end seismic imaging. Biondo is now Associate Professor (Research) of Geophysics and leads SEP efforts in 3-D imaging. He is a member of SEG and EAGE.



Morgan Brown received a B.A. in Computational and Applied Mathematics from Rice University in 1997 and is currently completing a Ph.D. at SEP. Morgan worked as a research intern with Western Geophysical in 1997 and with Landmark Graphics in 2000. He is a member of SEG.



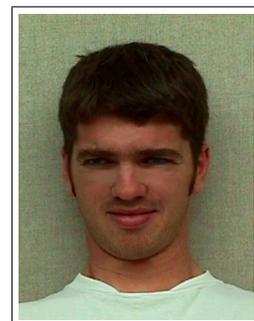
Jon F. Claerbout (M.I.T., B.S. physics, 1960; M.S. 1963; Ph.D. geophysics, 1967), professor at Stanford University, 1967. Best Presentation Award from the Society of Exploration Geophysicists (SEG) for his paper, *Extrapolation of Wave Fields*. Honorary member and SEG Fessenden Award “in recognition of his outstanding and original pioneering work in seismic wave analysis.” Founded the Stanford Exploration Project (SEP) in 1973. Elected Fellow of the American Geophysical Union. Authored three published books and five internet books. Elected to the National Academy of Engineering. Maurice Ewing Medal, SEG’s highest award. Honorary Member of the European Assn. of Geoscientists & Engineers (EAGE). EAGE’s highest recognition, the Erasmus Award.



Robert Clapp received his B.Sc.(Hons.) in Geophysical Engineering from Colorado School of Mines in May 1993. He joined SEP in September 1993, received his Masters in June 1995, and his Ph.D. in December 2000. He is a member of the SEG and AGU.



Sean Crawley Graduated from Arizona, joined SEP, and best of all got married in 1994. He received his Masters in 1996. His current research interests are interpolation and multiple suppression.



Sergey Fomel received his Diploma in Geophysics from Novosibirsk University in 1990. From 1990 to 1994 he worked at the Institute of Geophysics (Siberian Branch of the Russian Academy of Sciences) in Novosibirsk. He joined the Stanford Exploration Project in 1994 and received a Ph.D. in Geophysics in 2001. During six months of 1998 Sergey worked at Schlumberger/Geco-Prakla in England. He is a member of SEG, AGU, and SIAM.



Doug Gratwick received his B.S. degree in Geology from Colorado State University in May 1999. He joined SEP in September 1999 and is currently working towards a M.S. in Geophysics. His research projects currently include imaging complex elastic models and seismic interpretation for the SAGE program (Summer of Applied Geophysical Experience). He is a member of SEG and AAPG.



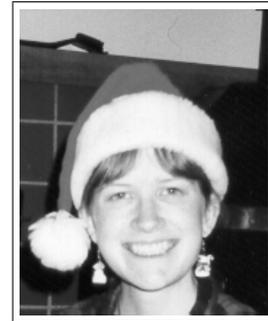
Antoine Guitton received a MSc in geophysics from Universite de Strasbourg, France in 1996 and from Stanford University in 2000. Received a "Diplome d'ingenieur de L'Ecole de Physique du Globe de Strasbourg" in 1996. He received the Best Student Paper Award from the SEG in 1999. Assistant research geophysicist at the Institut Francais du Petrole (Paris-1996/97) working on well seismic imaging. Assistant research geophysicist at CGG Houston (1997-98) working on multiples attenuation. He joined SEP in September 1998. Current research topics are nonlinear inversion and noise attenuation. He is member of the SEG.



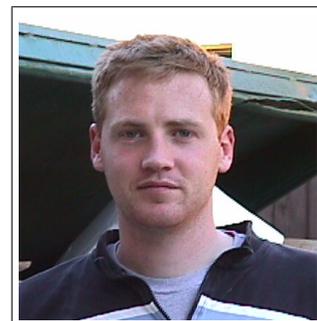
Carmen Mora received her B.S. in Computer Engineering (1986) and her M.S. in Computer Science (1992) from Universidad Simon Bolivar, Venezuela. From 1986 to 1994 she worked as a system developer and system administrator in UNIX platforms. She has been employed by PDVSA INTEVEP S.A. since 1993, working mainly in the development of scientific visualization tools for seismic data. In January 1999, she joined SEP and she received her M.S. degree in Geophysics in December 2000. She is currently working at PDVSA INTEVEP S.A.



Marie Prucha received her B.Sc.(Hons.) in Geophysical Engineering from Colorado School of Mines in May 1997. She joined SEP in September 1997 and received her MS in June 1999. She is currently working towards a Ph.D. in geophysics. She is a member of SEG.



James Rickett graduated with a B.A. in Natural Science from Cambridge University in 1994, and obtained an M.Sc. in Exploration Geophysics from Leeds University in 1995. He is currently working towards a Ph.D. in geophysics at Stanford, and is a member of the SEG and the EAGE.



Daniel A. Rosales received his BS in Geophysics from Universidad Simon Bolivar, Venezuela, in 1997. From 1996 to 1998 he worked at PDVSA-NTEVEP S.A. in seismic modeling, 2-D prestack depth migration, and velocity estimation by wave-field downward continuation. In 1998 he joined Simon Bolivar University as a instructor in geophysics and received a scholarship from PDVSA-CIED. In september 1999, he became a new member of the Stanford Exploration Project and he is currently working towards a M.Sc. in geophysics. He is a member of SEG and AAPG.



Paul Sava graduated in June 1995 from the University of Bucharest, with an Engineering Degree in Geophysics. Between 1995 and 1997 he was employed by Schlumberger GeoQuest. He joined SEP in 1997, received his M.Sc. in 1998, and continues his work toward a Ph.D. in Geophysics. His main research interest is in seismic imaging using wave-equation techniques. He is a member of SEG, EAGE and the Romanian Society of Geophysics.



Louis Vaillant graduated from the Ecole des Mines de Paris (M.S.) and joined SEP in December 1998 for his military obligation, supported by Elf Exploration Production. He spent one summer with Elf UK at the Geoscience Research Centre in London (1997), and worked on earthquake rupture processes at the Ecole Normale Superieure de Paris (1998). Louis' current research is on wave-equation migration. He is a member of SEG.



Sponsors of the Stanford Exploration Project, 1999-2000