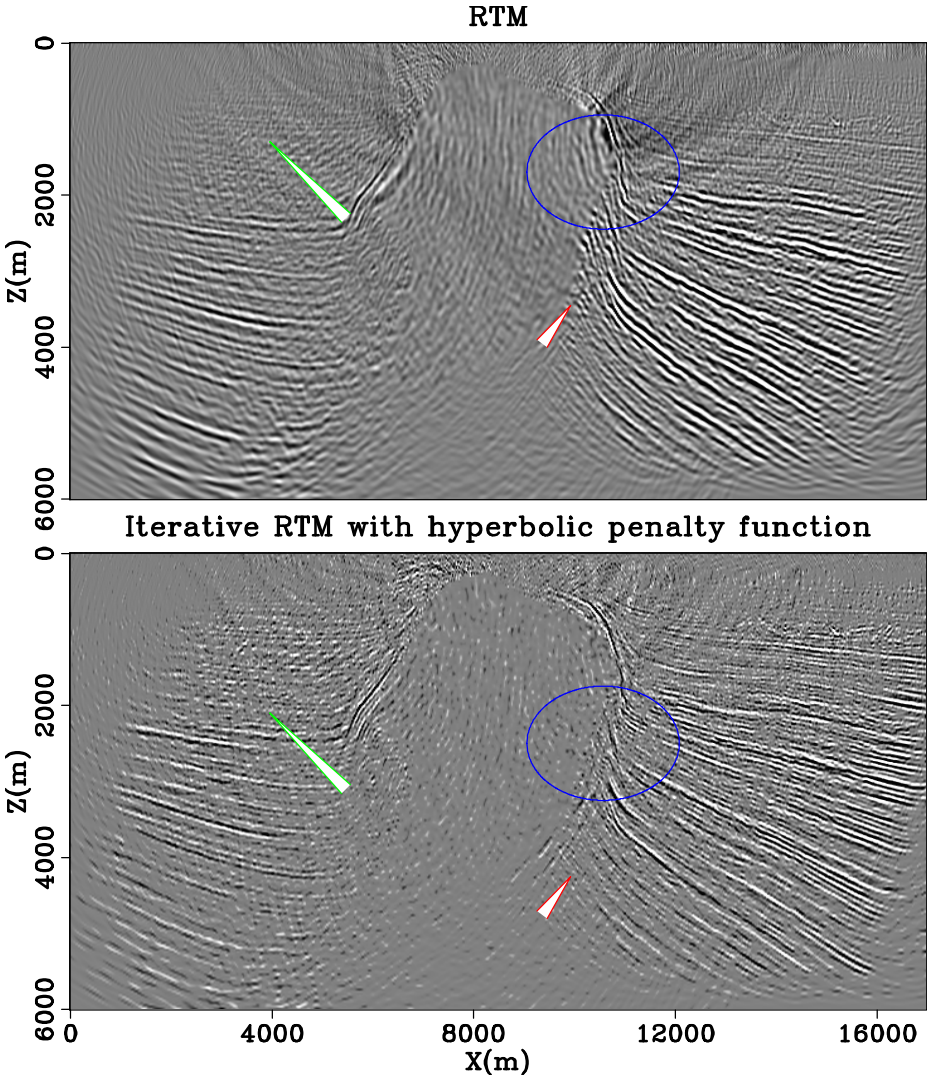
STANFORD EXPLORATION PROJECT

Ali Almomin, Gustavo Alves, Ohad Barak, Guillaume Barnier, Biondo Biondi, Ettore Biondi, Robert Brune, Jon Claerbout, Robert Clapp, Taylor Dahlke, Carlo Fortini, Antoine Guitton, Chris Leader, Stewart Levin, Yinbin Ma, Musa Maharramov, Paul Milligan, Sjoerd de Ridder, Shuki Ronen, Xukai Shen, and Mandy Wong

Report Number 155, October 2014



Preface

The electronic version of this report[1] 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 that allows both the author and the reader of a publication to verify the reported results. Reproducibility facilitates the transfer of knowledge within SEP and between SEP and its sponsors.

ER denotes Easily Reproducible and are the results of 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, be easily available to all researchers (e.g., SEG-EAGE data sets), or be available in the SEP data library[2]. We assume you have a UNIX workstation with Fortran, Fortran90, C, C++, X-Windows system and the software downloadable from our website (SEP makerules, SEPlib, and the SEP latex package), or other free software such as SU. 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. The primary reasons for the CR designation is that the processing requires 20 minutes or more, MPI or CUDA based code, or commercial packages such as Matlab or Mathematica.

NR denotes Non-Reproducible figures. SEP discourages authors from flagging their figures as NR except for figures that are used solely for motivation, comparison, or illustration of the theory, such as: artist drawings, scannings, or figures taken from SEP reports not by the authors or from non-SEP publications.

Our testing is currently limited to LINUX 2.6 (using the Intel compiler), but the code should be portable to other architectures. Reader's suggestions are welcome. More information on reproducing SEP's electronic documents is available online[3].

[1] http://sepwww.stanford.edu/private/docs/sep155
[2] http://sepwww.stanford.edu/public/docs/sepdatalib/toc_html
[3] http://sepwww.stanford.edu/research/redoc/

SEP155 — TABLE OF CONTENTS

Imaging and inversion

Case studies

Signal processing and L1/L2 optimization

Modeling and anisotropy

Velocity estimation

Image space separation of linearly blended data

Chris Leader and Biondo Biondi

ABSTRACT

Separating simultaneously acquired seismic data is the link between more efficient acquisition and conventional imaging techniques. Successful methods of separating these data rely strongly on random source timings and positionings; loosening this acquisition restriction would make survey design and implementation more flexible. By performing a series of transformations it is possible to isolate and remove overlapping artifacts that are ubiquitious when imaging simultaneously acquired data with constant time delays. Initially, Extended Reverse Time Migration is applied, generating a roughly focused image in subsurface offset. The image is transformed to the angle domain and a hyperbolic radon transform is then applied, isolating certain events and allowing a separation to be performed as a function of curvature. Basic tests have shown that after a few iterations of this transform events become well separated. The reverse transforms are then applied and the data demigrated, giving the equivalently unblended dataset without requiring accurate velocity control.

INTRODUCTION

Contemporary seismic targets are increasingly often associated with steeply dipping structures and strong velocity contrasts. In order to illuminate these difficult features data with large offsets and multiple source boats are acquired (Verwest and Lin, 2007). Intuitively this leads to both more expensive acquisition and an increase in field waiting time. This latter ramification is due to the fact that it is necessary to allow the energy from the previous source to sufficiently dissipate before recording the next source point. If waiting time was not a restriction then denser sampling could be recorded per unit time and acquisition would be significantly more efficient (Beasley (2008); Hampson (2008); Berkhout and Blacquiere (2008)). Practically, it is possible to disregard this waiting time and fire the next shot when in position; this is often called continuous recording of seismic data. Recording overlapping data in this manner will require more processing time than conventionally acquired data, since separation will be necessary to mitigate imaging artifacts. However the economic gains from reduced acquisition time far outweigh this extra processing cost.

These simultaneously acquired data can be used to directly invert for model properties (Dai and Schuster (2009); Tang and Biondi (2009)). However such methods require exact velocity model knowledge. Separation and subsequent imaging could be integrated into production data flows; successful existing methods rely on random sampling in the source timings and locations (Abma and Yan (2009); Moore et al. (2008)). For example, constant receiver gathers can be transformed into the f-k or tau-p domain and iteratively thresholded (Doulgeris et al., 2011), iteratively removed in the parabolic random domain (Ayeni et al., 2011), removed by using a convex projection approach (Abma and Foster, 2010), or through compressive sensing methods (Herrmann et al., 2009).

Image domain processing has been used effectively for coherent energy removal/attenuation by posing the problem in the extended image space (Zhang et al. (2012); Sava and Guitton (2005)). It is possible to untangle certain events in this domain and recreate cleaner shot gathers by virtue of higher signal-to-noise ratio and reduced dimensionality. Additionally, when using the extended image space (Sava and Vasconcelos, 2011), event kinematics are preserved. Consequently, if the velocity model is inaccurate then demigration is still possible (Chauris and Benjemaa, 2010).

Similarly, blended data can be untangled by using extended imaging. This is done by imaging these data across a range of subsurface offsets and then applying a transformation to the angle domain. It is possible to distinguish events from separate shots in this domain, by then applying some variety of curvature dependent filtering energy from overlapping sources can be removed. These data can then be demigrated, resulting in the equivalently unblended dataset.

This discussion will investigate how these overlapping data manifest themselves in the Angle Domain Common Image Gathers (ADCIGs) and how best to isolate and remove the energy identified as noise. Subsurface offset domain image gathers will be constructed and compared for conventional data and data blended using a constant time delay, these will also be contrasted in the angle domain. By using a hyperbolic radon inversion methodology the energy from interfering shots can be focused, isolated and removed. Thus only events of interest will be left and these data can be demigrated, resulting in successful separation of blended data.

A number of existing techniques are successful in separating data acquired using a random timing approach, thus the focus of this study will be on data acquired using a constant time delay. This is referred to from hereon as linearly blended data.

ADCIG CONSTRUCTION

The imaging technique used for this study will be Reverse Time Migration (RTM.) This is an algorithm based on direct solutions of the wave equation, meaning that energy associated with multiply scattered events, steep dips and a broad range of wavenumbers will be preserved. This process can be described in equation 1.

$$m(\mathbf{x}) = \sum_{\mathbf{x}_s,\omega} f(\omega) G_0(\mathbf{x}, \mathbf{x}_s, \omega) \sum_{\mathbf{x}_r} G_0(\mathbf{x}, \mathbf{x}_r, \omega) d^*(\mathbf{x}_r, \mathbf{x}_s, \omega), \tag{1}$$

where $\mathbf{x}$ represents the spatial coordinates, $m(\mathbf{x})$ the scattering field, $\mathbf{x}_s$ the current source coordinates, $\mathbf{x}_r$ the current receiver coordinates, ω the temporal frequency, $d^*(\mathbf{x}_r, \mathbf{x}_s, \omega)$ the complex conjugate of the data and G_0 the relevant Green's function. Only the zero-offset image (Claerbout, 1971) is calculated and this will contain all necessary amplitude and kinematic information for demigration, assuming the velocity model accurately represents the data.

However, for the problem of separating continuously acquired data a stringent requirement on the velocity model is undesirable. Direct application of equation 1 with an incorrect velocity model will result in the loss of certain events, and subsequent demigration will not well represent the original dataset. To preserve all event kinematics extended imaging must

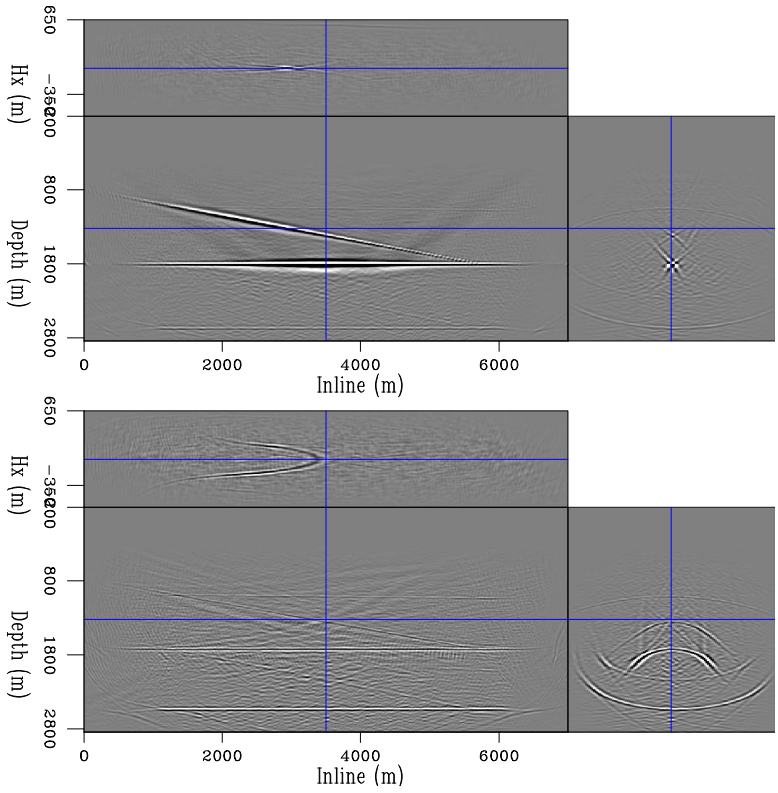


Figure 1: Linearly blended data migrated into the subsurface offset domain using the correct velocity model (top) and a model 10% too slow (bottom.) [**CR**] chris1/. offimgs

be used. If zero-offset imaging can be described by equation 2, then extended imaging can be described by equation 3.

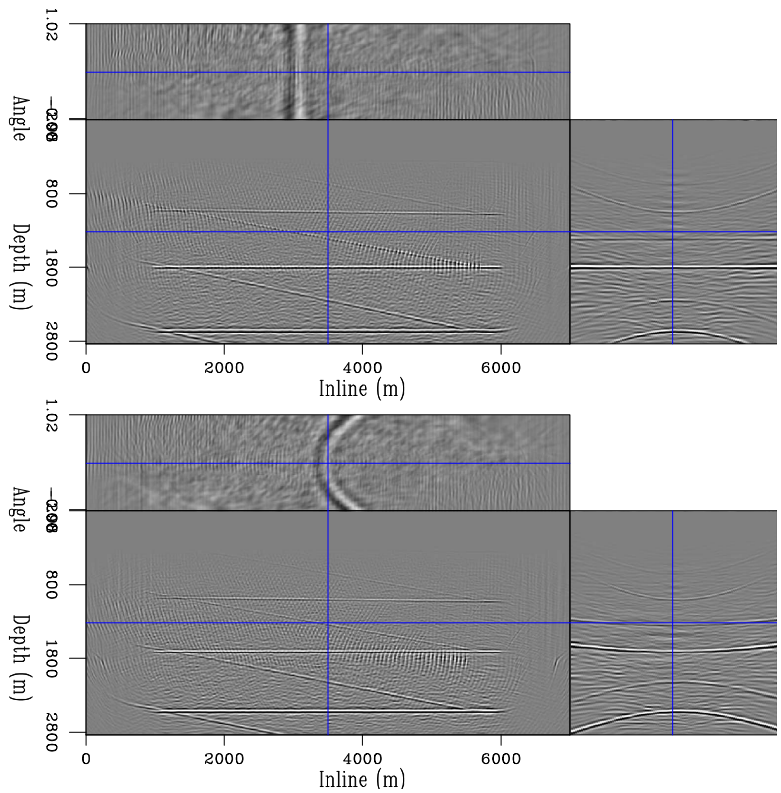


Figure 2: The same images as Figure 1 but with the third axis transformed into the subsurface angle domain, rather than subsurface offset. [**CR**] chris1/. angimgs

$$I(x,y,z) = \sum_{i}^{nshots} \sum_{t} P_s(x,y,z,t;\mathbf{s}_i) P_r(x,y,z,t;\mathbf{s}_i). \tag{2}$$

$$I(x,y,z,x_h,y_h) = \sum_{i}^{nshots} \sum_{t} P_s(x+x_h,y+y_h,z,t;\mathbf{s}_i) * P_r(x-x_h,y-y_h,z,t;\mathbf{s}_i) \tag{3}$$

Here, $I(x,y,z)$ is the image in space, P_s is the source wavefield and P_r is the receiver wavefield. If lag coordinates in x and y are introduced (x_h and y_h), a 5D image can be created. It is possible to have lags in both t and z to create a 7D image, or any combination thereof. From here on this discussion will be limited to subsurface offsets in the x direction only.

If the correct velocity model was used for imaging then the energy will be focused to a point in subsurface offset. If an incorrect model was used then the energy will be spread out over a range of offsets. Analysing this moveout as a function of the velocity model is the

core concept of Wave Equation Migration Velocity Analysis (WEMVA) (Sava and Biondi, 2003).

Figure 1 shows an image of some simple linearly blended data imaged with both the correct velocity model and with an incorrect model. In both cases the energy resulting from the overlapping data are readily identifiable. To filter this overlapping energy it is desirable to apply a second transform to these offset panels that can isolate events according to their curvature. In this paper a hyperbolic radon transform will be used.

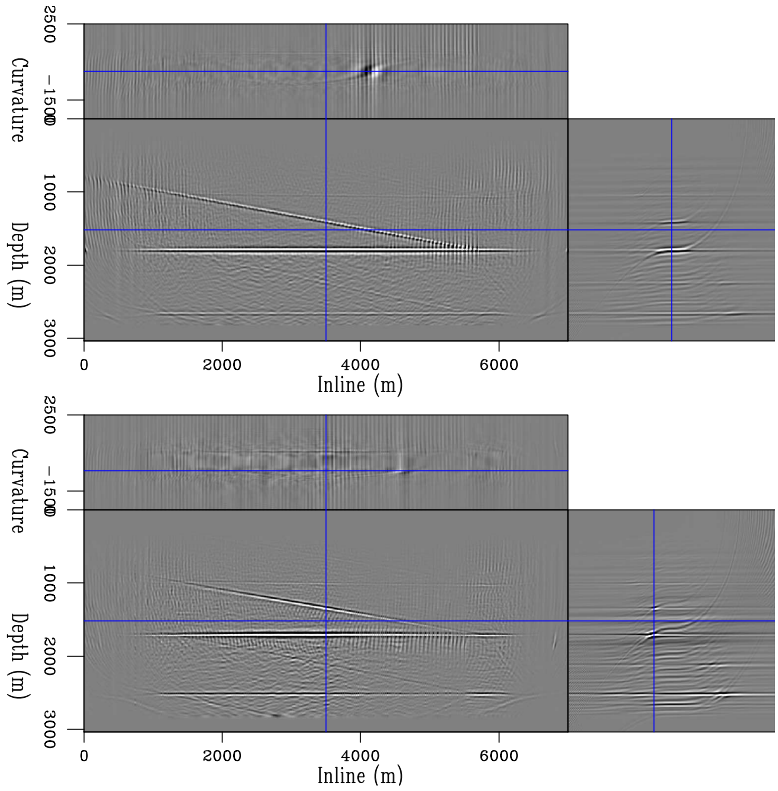


Figure 3: The angle domain image with the correct velocity model after a single hyperbolic transform (top) and the slow velocity image after a single hyperbolic radon transform. [**CR**] chris1/. anghypcomp

Initial inspection shows that the subsurface offset domain is not the ideal domain for this moveout-based separation. If a radon transform is applied to a focused point, the result will be a line in the transform space. We would like our data to either be all points or all lines, not a mixture of the two. While in this case the incorrect velocity model only results in curves it is not desirable to restrict the case where the velocity may be accurate for certain events. A simple transform from the offset domain to the angle domain can be used. Figure 2 shows the same two panels as in Figure 1, but in the angle domain (ADCIGs).

It is still possible to clearly identify the events from the primary data and overlapping data. However, now all the events in this new domain are associated with a given curvature and applying a curvature based transform will not result in the spreading out of some of this energy, but rather focusing of all events.

RADON INVERSION IN THE IMAGE SPACE

Initially, a hyperbolic radon transform was applied to the ADCIGs in Figure 2, and the result can be seen in Figure 3. Two things are evident from studying these images - the transform has been successful in beginning to separate these data according to their moveout, however they are not well focused in this new domain. There is some clear transform noise and a spreading over curvature values for all events. To improve this focusing a simple gradient based inversion can be used instead of a direct adjoint methodology. By using this transform, it's adjoint and a conjugate directions solver this gradient based inversion can be easily constructed (Claerbout, 2001).

Figure 4 shows the result of applying one instance of this radon transform and the result after applying ten iterations of this radon based inversion to the image obtained with the correct velocity model. It is immediately clear that the latter of these served to greatly improve the focusing of these events in curvature space, and that curvature based filtering will remove the overlapping data. The reverse transforms can then be applied and the separated data constructed. Figure 5 then shows these two panels but for the image obtained with a velocity model which was 10% too slow.

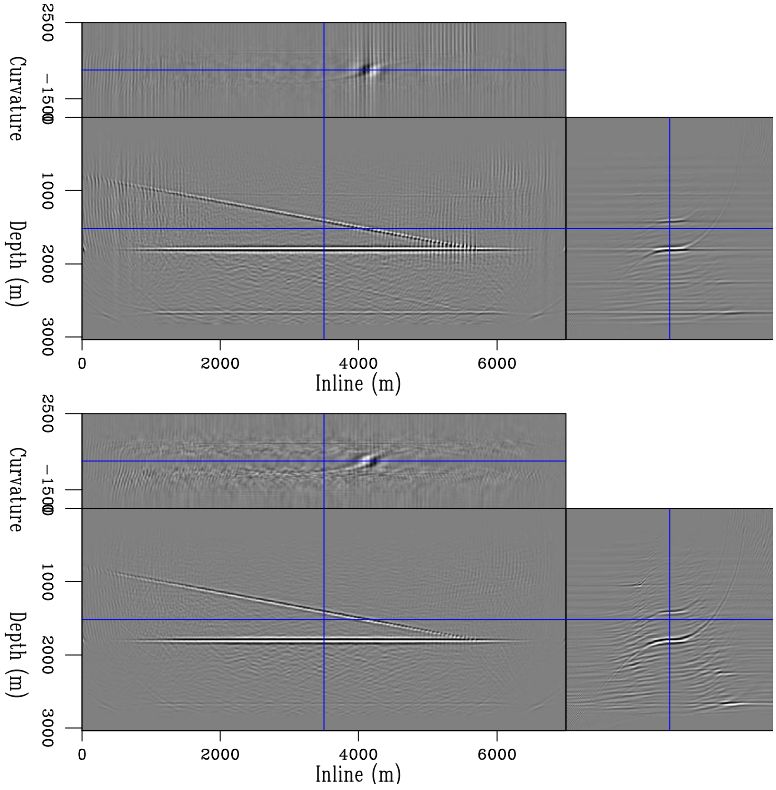


Figure 4: The angle domain image with the correct velocity model after a single hyperbolic transform (top) and after ten linear iterations of the transform (bottom.) Note the focusing at zero curvature (corresponding to a flat angle gather.) [**CR**] chris1/. anghyp

Again, the inversion was able to more tightly focus these events for given values of curvature. This focusing now occurs at a non-zero value of curvature, however the relative differences in curvature values between primary and overlapping data are the same. Thus

these events can still be filtered according to their curvature, even though these data were imaged with an incorrect velocity model. The events do not focus at an exact curvature value because a hyperbola does not describe these shapes exactly. However, within a range of curvature values the overlapping events are still localised, meaning penalisation or filtering can be used.

CONCLUSIONS

It is demonstrated herein that it is possible to untangle linearly blended data in the image domain. Migrating these data into ADCIGs serves as the first step, resulting in a space where it is possible to distinguish between primary data and overlapping data. A hyperbolic radon inversion can then isolate these events into given measurements of curvature; certain curvature values can then be removed or penalised, effectively removing energy associated with overlapping shots. The remaining energy can then be transformed back to ADCIGs and then demigrated, with the final output being the separated dataset. This separation scheme does not depend on a well constratined velocity model.

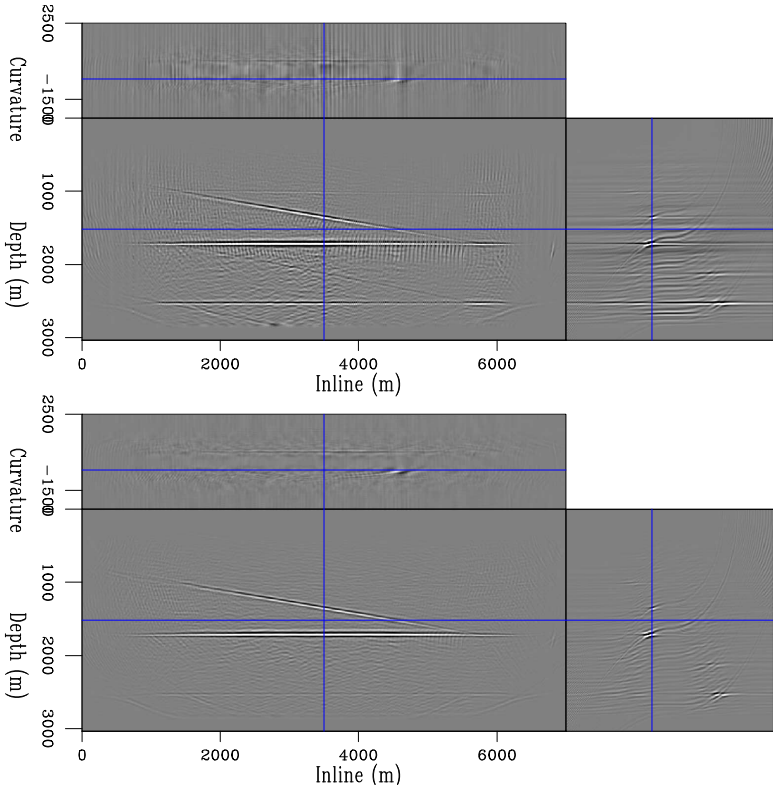


Figure 5: The angle domain image with the incorrect velocity model after a single hyperbolic transform (top) and after ten linear iterations of the transform (bottom), note the focusing is now at non-zero curvature. [**CR**] chris1/. slowhyp

FUTURE WORK

While the hyperbolic transform does focus these events over a smaller range of curvature values, the events in the angle domain can not be exactly described by hyperbolae. This is even more true for data acquired over a more complex model. It would be better to decompose these events into parabolic curvature and tangent values, this will be attempted next.

REFERENCES

Abma, R., M. T. A. M. Y. J. and M. Foster, 2010, High quality separation of simultaneous sources by sparse inversion: 72nd EAGE Conference and Exhibition Abstracts, **72nd**.

Abma, R. and J. Yan, 2009, Separating simultaneous sources by inversion: 71st EAGE Conference and Exhibition Abstracts, **71st**.

Ayeni, G., A. Almomin, and D. Nichols, 2011, On the separation of simultaneous source data by inversion, chapter 4, 20–25.

Beasley, C. J., 2008, A new look at marine simultaneous sources: The Leading Edge, **27**.

Berkhout, A., V. D. and G. Blacquiere, 2008, Processing of blended seismic data: SEG Technical Program Expanded Abstracts, **78th**.

Chauris, H. and M. Benjemaa, 2010, Seismic wave-equation demigration/migration: GEOPHYSICS, **75**, S111–S119.

Claerbout, J., 2001, Image estimation by example: Society of Exploration Geophysicists.

Claerbout, J. F., 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, 467–481.

Dai, W. and J. Schuster, 2009, Least-squares migration of simultaneous sources data with a deblurring filter: SEG Expanded Abstracts, **79th**, 2990–2993.

Doulgeris, P., A. Mahdad, and G. Blacquiere, 2011, Iterative separation of blended marine data: Discussion on the coherencepass filter, chapter 5, 26–31.

Hampson, G., S. J. H. F., 2008, Acquisition using simultaneous sources: SEG Technical Program Expanded Abstracts, **78th**.

Herrmann, F., Y. Erlangga, and T. Lin, 2009, Compressive simultaneous full-waveform simulation: GEOPHYSICS, **74**, A35–A40.

Moore, I., B. Dragoset, T. Ommundsen, D. Wilson, C. Ward, and D. Eke, 2008, Simultaneous source separation using dithered sources, chapter 565, 2806–2810.

Sava, P. and B. Biondi, 2003, Wave equation migration velocity analysis by inversion of differential image perturbations, chapter 540, 2124–2127.

Sava, P. and A. Guitton, 2005, Multiple attenuation in the image space: Geophysics, **70**, V10–V20.

Sava, P. and I. Vasconcelos, 2011, Extended imaging conditions for wave-equation migration: Geophysical Prospecting, **59**, 35–55.

Tang, Y. and B. Biondi, 2009, Least square migration/inversion of blended data: SEG Expanded Abstracts, **28**, 2859–2863.

Verwest, B. and D. Lin, 2007, Modelling the impact of wide-azimuth acquisition on subsalt imaging: Geophysics, **72**, 241–250.

Zhang, Y., L. Duan, et al., 2012, Predicting multiples using a reverse time demigration: Presented at the 2012 SEG Annual Meeting.

Chaotic wavefield gradiometry

Sjoerd de Ridder and Biondo Biondi

ABSTRACT

We propose a new technique for passive seismic imaging via the direct application of operators to noise recordings. We propose a time-domain 2D scalar wave equation to describe the propagation of surface waves within a narrow frequency range. Outside the source region, this scalar wave equation relates the second-order spatial and second-order temporal derivatives of the wavefield with the local velocity. Different from seismic interferometry, this technique does not rely on cross-correlations to reveal the statistical coherence of a chaotic wavefield at two locations. Rather, it relies on the local measurements of velocity obtained directly from the ratio between temporal and spatial derivatives of the wavefield. The new method allows us to do passive imaging with much shorter passive recordings. Numerical data examples show that this theory can yield reliable images if the wavefield is sampled sufficiently in space and time.

INTRODUCTION

Geophysicists have long attempted to image chaotic wavefields. Aki (1957) first derived the dispersion of surface waves from the cross-correlations of a circle of stations. Claerbout (1968) showed that the one dimensional auto-correlation of transmission responses would yield the reflection response and later conjectured an extension to three-dimensions by cross-correlations. See Wapenaar et al. (2008) for a good review of seismic interferometry. Almost all approaches have centered on correlating long ambient seismic recordings made at two stations. Seismic interferometry has found its most wide application in the retrieval of surface waves (Campillo and Paul, 2003; Shapiro and Campillo, 2004; Shapiro et al., 2005; Gerstoft et al., 2006; Yao et al., 2006; Lin et al., 2008).

However, the advent of ever larger and denser arrays covering the earth's surface provides a complete recording of the surface-wave wavefield not aliased in either time or space in many regions and frequency bands of interest. This essentially entails a direct measurement of local medium properties because the temporal and spatial derivatives of the wavefield are related by the medium properties through the wave equation. Here we exploit this relationship by the development of a wavefield gradiometry method that can be directly applied to chaotic (and non-chaotic) scalar wavefields.

Using wavefield gradients to infer propagation velocity is not a new concept. Using the gradients of the wavefield gradiometry to extract the velocity field has been proposed by Langston (2007a,c,b). The first-order spatial and temporal derivatives of the wavefield amplitudes can be inverted for a set of wavefield coefficients, which relate to the local ray parameter, local wave directionality, local geometrical spreading and local radiation pattern. This technique was applied on recordings of the Embayment Seismic Excitation Experiment in a one-dimensional linear array (Langston, 2007a,b), in two dimensions using

local earthquakes recorded by a small array near Moscow Tennessee (Langston, 2007c) and on several earthquakes recorded by US Array (Liang and Langston, 2009).

The fundamental assumption of this work is that the wavefield at each point consists of non-overlapping propagating plane waves from a point source (Langston, 2007a,c). This assumption limits the use of wavefield gradiometry to deterministic wavefields where specific arrivals can be identified. However, here we pose a two-dimensional scalar wave equation for the propagation of a chaotic wavefield consisting of single mode surface waves. We devise a strategy for chaotic wavefield gradiometry based on second-order partial derivatives of the observed chaotic wavefield amplitudes.

In this paper we first develop the method as a discrete computation for wavefields densely and irregularly sampled in space and regularly in time. Then I provide several numerical examples using both an analytic and finite-difference solution to the two-dimensional scalar wave equation.

CHAOTIC WAVEFIELD GRADIOMETRY

We assume that surface wave propagation in two dimensions is governed by a two-dimensional scalar wave equation for each frequency component of the wavefield:

$$\mathrm{c}^2(x,\omega)\nabla^2\hat{\mathrm{u}}(x,\omega) + \omega^2\hat{\mathrm{u}}(x,\omega) = -\hat{\mathrm{s}}(x,\omega), \tag{1}$$

where u is the wavefield variable observable in time and space and $\hat{\mathrm{u}}$ its Fourier transformed counterpart, s is a generalized source term and $\hat{\mathrm{s}}$ its Fourier transformed counterpart, ∇^2 is the Laplacian acting on the spatial dimensions and c is the phase velocity. After a sufficiently narrow bandpass filter (with central frequency ω'), for frequencies over which we can neglect the frequency dependence of the phase velocity, the filtered wavefield obeys:

$$\mathrm{c}^2_{\omega'}(x)\nabla^2\mathrm{u}(x,t) - \partial_t^2\mathrm{u}(x,t) = -\mathrm{s}(x,t), \tag{2}$$

where ∂_t^2 is the second-order derivative acting on the time dimension and $\mathrm{c}_{\omega'}$ is the phase velocity for this central frequency.

For chaotic wavefields the source distribution is generally unknown. Unless local sources dominate, we can assume that the source distribution is zero within the area of recordings:

$$\nabla^2\mathrm{u}(x,t)\ \mathrm{c}^2_{\omega'}(x) = \partial_t^2\mathrm{u}(x,t). \tag{3}$$

If sufficiently dense recordings are available we can evaluate the spatial and temporal second-order derivatives by finite differences. Let $\mathbf{u}(\mathbf{x},\mathbf{t})$ denote the discrete recordings of the wavefield, and $\mathbf{D_{tt}}$ and $\mathbf{D_{xx}}$ are (usually sparse) matrices containing finite-difference approximations of the second-order derivative operators applied to the wavefield in time and space, respectively. Then for each time slice $\mathbf{u}_i = \mathbf{u}(\mathbf{x}, t_i)$ equation 3 can be written as:

$$\mathbf{W}\mathrm{diag}\left\{\mathbf{D_{xx}}\mathbf{u}_i\right\}\ \mathbf{c}^2_{\omega'}(\mathbf{x}) = \mathbf{W}\mathbf{D_{tt}}\mathbf{u}_i, \tag{4}$$

where $\mathbf{D_t}$ operates on a few adjacent time slices, diag{ } denotes a diagonal operator specifying the elements on the diagonal between { }. We can discard locations with poor measurements using a masking operator, $\mathbf{W}$, which has the structure of a diagonal matrix with ones and zeros on the diagonal elements. Equation 4 has the structure $\mathbf{F}_i\,\mathbf{m} = \mathbf{b}_i$,

where $\mathbf{m} = \mathbf{c}^2_{\omega'}(\mathbf{x})$, $\mathbf{F}_i = \text{diag}\{\mathbf{D_{xx}u}_i\}$, and $\mathbf{b}_i = \mathbf{D_{tt}u}_i$. We form these data fitting equations for a set of N_t time slices in the data, and the least-squares estimator for $\mathbf{m}$ is given by inverting:

$$\sum_{i=1}^{N_t} \mathbf{F}_i^\dagger \mathbf{W}^\dagger \mathbf{W} \mathbf{F}_i \; \mathbf{m} = \sum_{i=1}^{N_t} \mathbf{F}_i^\dagger \mathbf{W}^\dagger \mathbf{W} \mathbf{b}_i, \tag{5}$$

where † denotes matrix adjoint. The solution for each element of $\mathbf{m}(\mathbf{x})$ is essentially an independent division of the second-order temporal derivative of the wavefield by the second-order spatial derivative of the wavefield. This ratio is evaluated at each point in space and averaged over all recording time by linear regression. We implicitly assume that the velocity does not change significantly across the length of the spatial finite-difference stencil.

We wish to add a spatial smoothness constraint on $\mathbf{m}$ by adding the spatial smoothness regularization $\epsilon \nabla^2 \mathbf{m} = \mathbf{0}$, where ϵ determines the importance of the model-styling goal versus the data-fitting equations. To aid the inversion, we decompose $\mathbf{m}$ into a background and a perturbation, $\mathbf{m} = \mathbf{m}_0 + \Delta\mathbf{m}$. This Laplacian is evaluated using the same finite-difference approximation $\mathbf{D_x}$ as before. The least-squares estimator for $\Delta\mathbf{m}$ is given by inverting:

$$\left[\sum_{i=1}^{N_t} \mathbf{F}_i^\dagger \mathbf{W}^\dagger \mathbf{W} \mathbf{F}_i + \epsilon^2 \mathbf{D}_{\mathbf{xx}}^\dagger \mathbf{D_{xx}}\right] \Delta\mathbf{m} = \tag{6}$$
$$\sum_{i=1}^{N_t} \mathbf{F}_i^\dagger \mathbf{W}^\dagger \mathbf{W} \mathbf{b}_i - \left[\sum_{i=1}^{N_t} \mathbf{F}_i^\dagger \mathbf{W}^\dagger \mathbf{W} \mathbf{F}_i + \epsilon^2 \mathbf{D}_{\mathbf{xx}}^\dagger \mathbf{D_{xx}}\right] \mathbf{m}_0.$$

We choose $\mathbf{m}_0$ to be a scalar and an estimate for the background model obtained from the ratio between the second-order temporal derivative and second-order spatial derivative averaged over all time and all points in space:

$$\text{m}_0 = \frac{1}{X} \sum_{j=1}^{N_s} \text{diag}\left\{\sum_{i=1}^{N_t} \mathbf{F}_i^\dagger \mathbf{W}^\dagger \mathbf{W} \mathbf{b}_i\right\}_j^{-1} \text{diag}\left\{\sum_{i=1}^{N_t} \mathbf{F}_i^\dagger \mathbf{W}^\dagger \mathbf{W} \mathbf{F}_i\right\}_j, \tag{7}$$

where j is the index over the elements of the diagonals, N_s is the total number of stations, and X is the cardinality of the set in the masking matrix, $\mathbf{W}$. Because $\mathbf{m}_0 = \text{m}_0$ is a constant over space, its spatial second-order derivative vanishes, and equation 6 reduces to:

$$\left[\sum_{i=1}^{N_t} \mathbf{F}_i^\dagger \mathbf{W}^\dagger \mathbf{W} \mathbf{F}_i + \epsilon^2 \mathbf{D}_{\mathbf{xx}}^\dagger \mathbf{D_{xx}}\right] \Delta\mathbf{m} = \sum_{i=1}^{N_t} \mathbf{F}_i^\dagger \mathbf{W}^\dagger \mathbf{W} \mathbf{b}_i - \sum_{i=1}^{N_t} \mathbf{F}_i^\dagger \mathbf{W}^\dagger \mathbf{W} \mathbf{F}_i \; \mathbf{m}_0. \tag{8}$$

Finally we retrieve the phase velocity, $c(x)$, from the inverted model perturbation using $c(x) = \sqrt{\text{m}_0 + \Delta\text{m}(x)}$. This method for imaging a chaotic wavefield does not employ cross-correlations.

NUMERICAL DATA EXAMPLES

We test the method presented in the previous section on three synthetic models: The first model is a homogeneous medium with velocity 2000 m/s. The second model is a Gaussian

anomaly of 50 m/s on a homogeneous background medium with velocity 2000 m/s. The third model is a two layer medium with 1950 m/s and 2050 m/s. We consider three source distributions: The first distribution is one source in the center of the domain. The second distribution consists of two sources at either side of the center of the domain. The third distribution consists of a large number of sources acting randomly in time, and positioned randomly on a boundary outside the domain. The matrices $\mathbf{D_{tt}}$ and $\mathbf{D_{xx}}$ contain second order finite-difference approximations to the second order spatial and temporal derivatives.

For the first set of examples, I use a far-field analytic solution in a homogeneous medium. I compute the wavefield radiating from a source located in the center of the medium. The wavefield is computed at a spacing of 0.5 m and with a sampling of 0.25 ms for a Ricker wavelet with central frequency of 75 Hz as the source. Because there is no noise or error, we omit the spatial smoothness regularization. The recovered velocity models for the three source distributions in a homogeneous medium are shown in Figure 1. Figure 2 shows profiles through the true and recovered homogeneous velocity models. The dashed lines in Figure 1 correspond to the locations of the profiles of Figure 2. Figures 1a-b and 2a-b reveal that the recovered velocity is very wrong near the sources. This is because the Green's solution used to model the wavefield is only a far-field solution and not a full solution to equation 2. When the sources are placed outside the domain to be imaged, we retrieve the correct velocity profile (Figure 2c). We computed the wavefield with very small temporal and spatial sampling intervals, so the error in the finite-difference approximation of the derivative operators is small. However, there is still a small discrepancy between the true and recovered velocities. This is due to the small remaining error in the finite-difference coefficients.

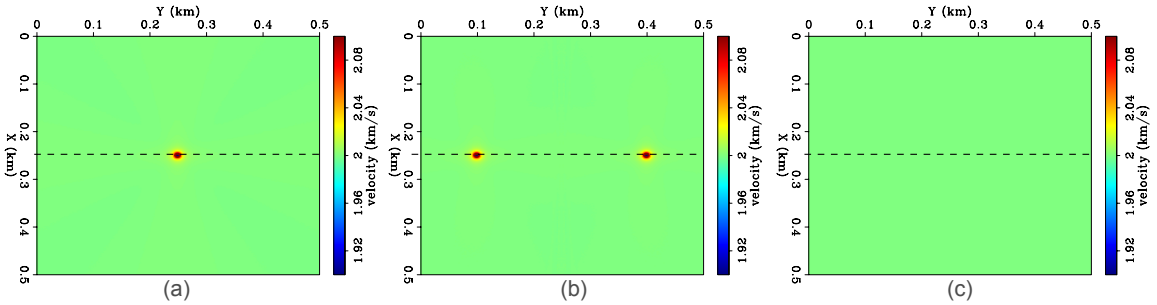


Figure 1: Recovered velocity models using an analytic solution to the wave equation in a homogeneous medium. a) Singe source at the center; b) Two sources on either side of the center; c) Sources placed randomly at the boundaries. Dashed lines indicate the locations of the profiles in Figure 2. [**ER**] sjoerd1/. anvels

The second set of examples is made using wavefields computed using finite differences (4th order in time and 16th order in space). For the three models and the three source distributions, the wavefields are computed at a spacing of 0.5 m and with a sampling of 0.25 ms for a Ricker wavelet with central frequency of 75 Hz as the source. For all three models and three source distributions the recovered models are shown in Figure 3. Profiles through true and recovered velocity models are shown in Figure 4. The dashed lines in Figure 3 correspond to the locations of the profiles of Figure 4. There is no large discrepancy between true and recovered velocity model in the vicinity of the sources. This is because the finite-difference modeling code includes both far-field and near-field terms

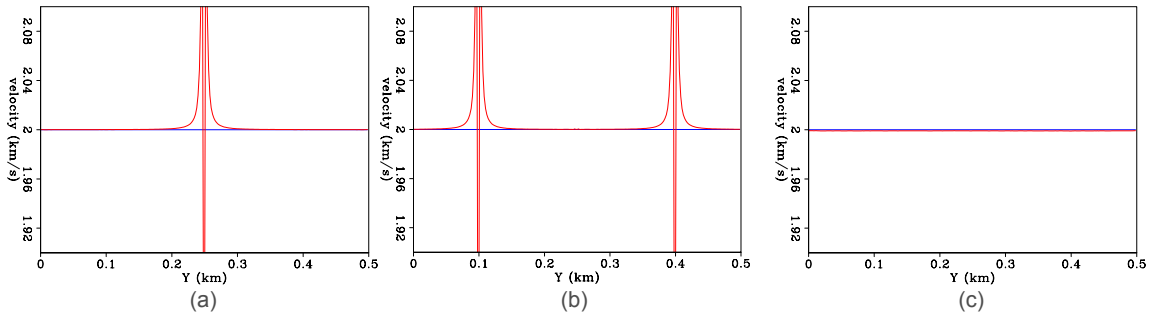


Figure 2: Profiles through true (blue) and recovered (red) homogeneous velocity models using an analytical solution to the wave equation in a: a) Single source at the center; b) Two sources on either side of the center; c) Sources placed randomly at the boundaries. [**ER**] sjoerd1/. anprofiles

of the solution to the wave equation. The distinct flower-like pattern centered around the source is due to the anisotropic error of the finite difference stencils in both the modeling step and the wavefield gradiometry step. One profile of recovered velocities in Figure 4 deviates significantly from the true velocity model (Figure 4h). This profile is taken exactly at a point in space where the wavefields from both sources interfere in time and the effective spatial wavelength becomes shorter. One explanation of the increased error may be that the error of the finite-difference stencil becomes significant.

In principle the method has no difficulty retrieving smooth or sharp velocity contrasts. However in practice, the resolution of the method cannot supersede the spatial sampling of the wavefield by the array. Furthermore, the resolution of chaotic wavefield gradiometry will depend on the length of the stencils used to evaluate the finite-difference approximations of the operators.

CONCLUSIONS

We presented a new method to extract surface-wave phase velocities from ambient seismic recordings: chaotic wavefield gradiometry. The method is based on evaluating the second-order spatial and temporal derivatives in the two-dimensional scalar wave equation to directly infer phase velocities. In contrast with conventional methods to image chaotic ambient seismic noise, this method does not employ cross-correlations. Numerical data examples show that the method would work well if the wavefield is recorded sufficiently in space and time. The retrieved velocity is very sensitive to the error in the approximations of the operators by the finite differences. The recovery of sharp velocity contrasts depends on sampling the wavefield very densely.

ACKNOWLEDGMENTS

We thank Joseph Stefani, Dimitri Bevc and Mark Meadows from Chevron's Center of Research Excellence (CORE) at Stanford and Gustavo Alves, Ohad Barack, Jason Chang, Bob Clapp, Jon Claerbout, Taylor Dahlke, Stew Levin, Musa Maharramov, Eileen Martin

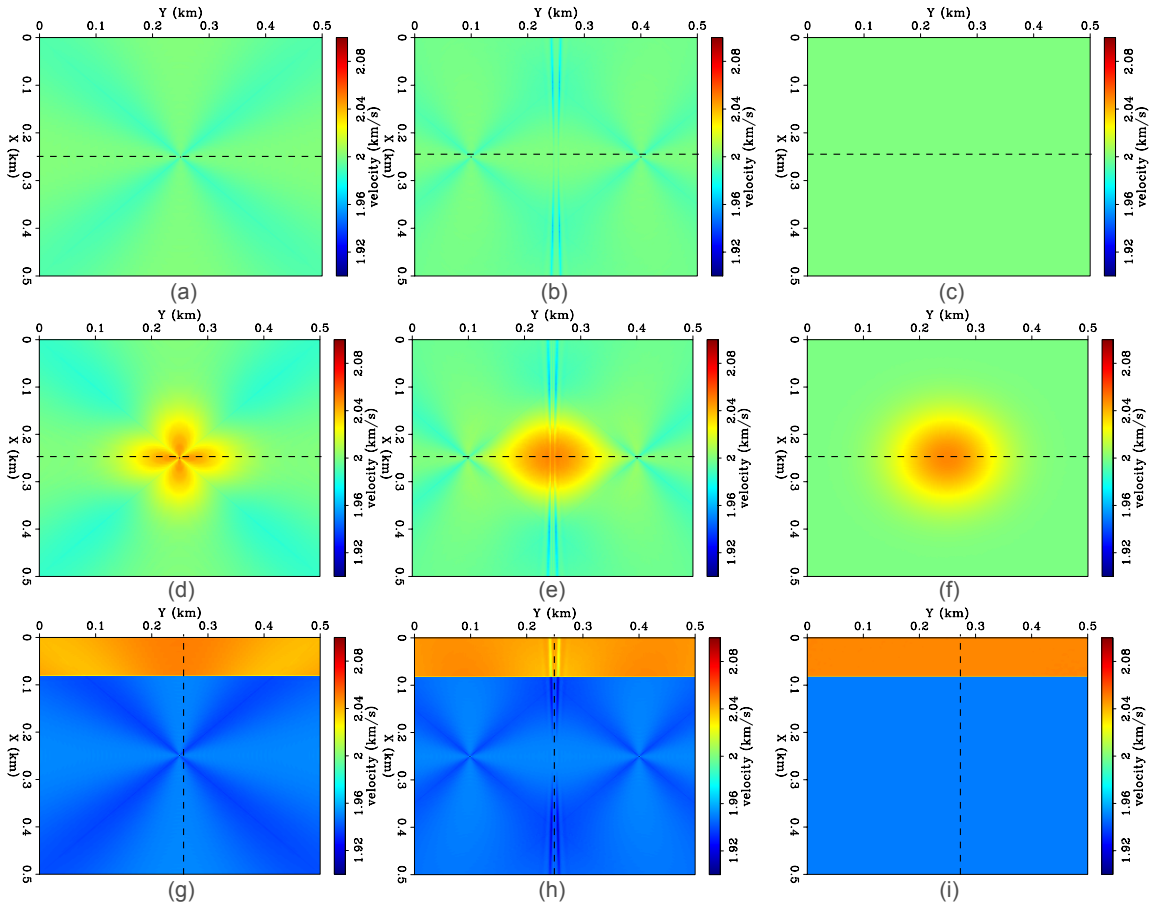


Figure 3: Recovered velocity models using finite difference solutions to the wave equation in: a-c) A homogeneous velocity model; d-f) A Gaussian anomaly; g-i) A two layer medium. Source distributions for the wavefields: a,d,g) Singe source at the center; b,e,h) Two sources on either side of the center; c,f,i) Sources placed randomly at the boundaries. Dashed lines indicate the locations of the profiles in Figure 4. [**ER**] sjoerd1/. vels

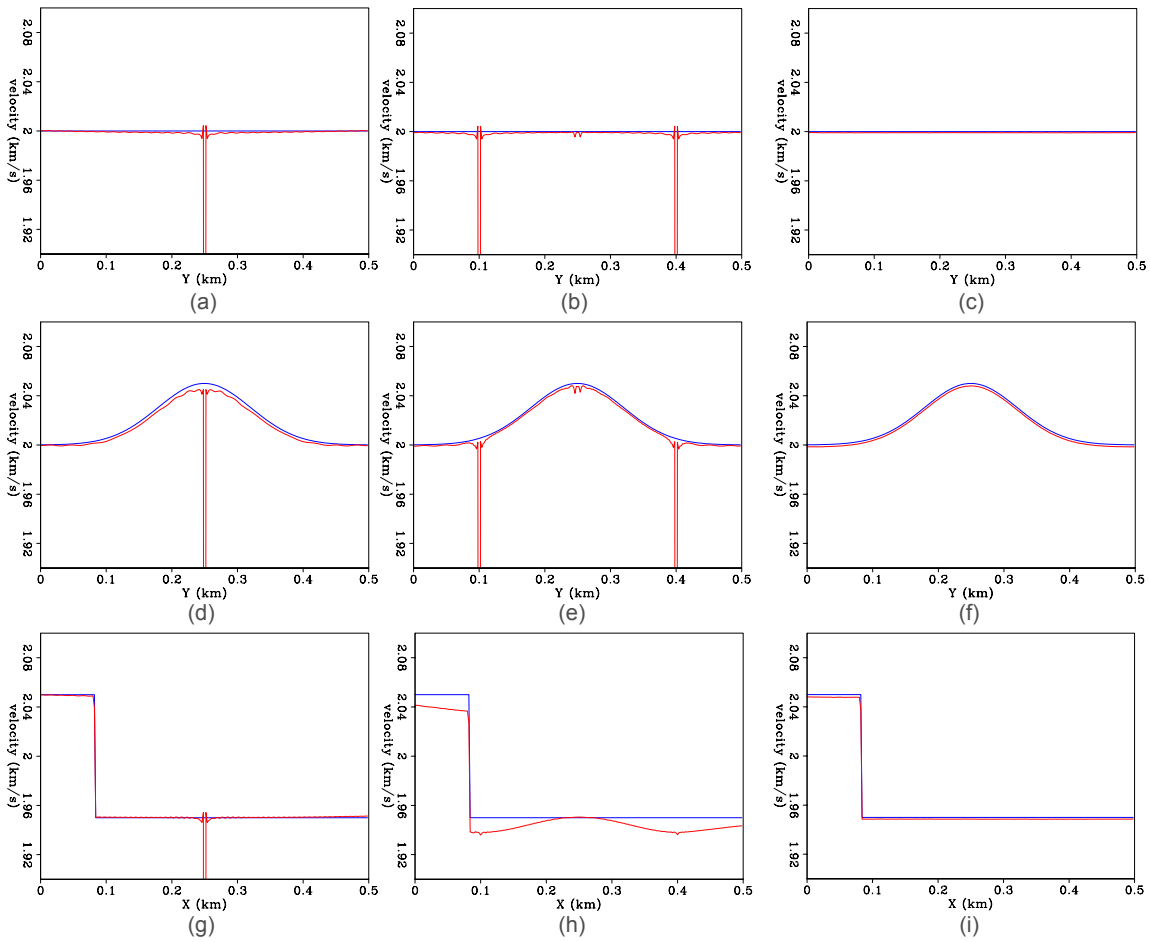


Figure 4: Profiles through true (blue) and recovered (red) velocity models using finite difference solutions to the wave equation in: a-c) A homogeneous velocity model; d-f) A Gaussian anomaly; g-i) A two layer medium. Source distributions for the wavefields: a,d,g) Singe source at the center; b,e,h) Two sources on either side of the center; c,f,i) Sources placed randomly at the boundaries. [**ER**] sjoerd1/. profiles

and Mandy Wong from the Stanford Exploration Project for helpful discussions and suggestions. Sjoerd de Ridder thanks Chevron's CORE at Stanford for supporting his visit to Stanford University during July and August of 2014.

REFERENCES

Aki, K., 1957, Space and time spectra of stationary stochastic waves, with special reference to microtremors: Bulletin of the Earthquake Research Institute, **35**, 415–456.

Campillo, M., and A. Paul, 2003, Long-range correlations in the diffuse seismic coda: Science, **299**, 547–549.

Claerbout, J. F., 1968, Synthesis of a layered medium from its acoustic transmission response: Geophysics, **33**, 264–269.

Gerstoft, P., K. G. Sabra, P. Roux, W. A. Kuperman, and M. C. Fehler, 2006, Green's functions extraction and surface-wave tomography from microseisms in southern California: Geophysics, **71**, no. 4, SI23–SI31.

Langston, C. A., 2007a, Spatial gradient analysis for linear seismic arrays: Bulletin of the Seismological Society of America, **97**, 265–280.

———, 2007b, Wave gradiometry in the time domain: Bulletin of the Seismological Society of America, **97**, 926–933.

———, 2007c, Wave gradiometry in two dimensions: Bulletin of the Seismological Society of America, **97**, 401–416.

Liang, C., and C. A. Langston, 2009, Wave gradiometry for usarray: Rayleigh waves: Journal of Geophysical Research: Solid Earth, **114**, n/a–n/a.

Lin, F.-C., M. P. Moschetti, and M. H. Ritzwoller, 2008, Surface wave tomography of the western united states from ambient seismic noise: Rayleigh and love wave phase velocity maps: Geophysical Journal International, **173**, 281–298.

Shapiro, N. M., and M. Campillo, 2004, Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise: Geophysical Research Letters, **31**, L07614–1–L07614–4.

Shapiro, N. M., M. Campillo, L. Stehly, and M. H. Ritzwoller, 2005, High-resolution surface-wave tomography from ambient seismic noise: Science, **307**, 1615–1618.

Wapenaar, K., D. Draganov, and J. Robertsson, 2008, Seismic interferometry: History and present status: Society of exploration geophysicists, geophysics reprint series no. 26: Society of Exploration Geophysicists.

Yao, H., R. D. Van Der Hilst, and M. V. De Hoop, 2006, Surface-wave array tomography in SE Tibet from ambient seismic noise and two-station analysis - I. Phase velocity maps: Geophysical Journal International, **166**, 732–744.

Improved depth imaging by constrained full-waveform inversion

Musa Maharramov and Biondo Biondi

ABSTRACT

We propose a formulation of full-wavefield inversion (FWI) as a constrained optimization problem, and describe a computationally efficient technique for solving *constrained full-wavefield inversion* (CFWI). The technique is based on using a total-variation regularization method, with the regularization weighted in favor of constraining deeper subsurface model sections. The method helps to promote "edge-preserving" blocky model inversion where fitting the seismic data alone fails to adequately constrain the model. The method is demonstrated on synthetic datasets with added noise, and is shown to enhance the sharpness of the inverted model and correctly reposition mispositioned reflectors by better constraining the velocity model at depth.

INTRODUCTION

Full-waveform inversion can achieve high resolution of subsurface velocity reconstruction where the target is shallow and well illuminated by refracted seismic energy from wide-offset surveys (Sirgue et al., 2010). However, illumination of deeper targets with refracted energy may require extra wide offset survey acquisitions, or otherwise suffer from poor constraining of deeper model sections. Mathematically, this is a manifestation of the fact that the full-waveform inversion is a mixed-determined problem, with shallow areas of the model overdetermined by the abundance of data, and deeper areas affected by poor resolution and spurious positioning errors. While geological priors such as well tie-ins may provide useful additional constraints in areas of poor illumination, regularization of the nonlinear inversion problems arising in full-waveform inversion is a well-established mathematical technique for dealing with underdetermined problems and noisy data. In particular, it is recognized that the *total variation* (TV) regularization promotes sparsity of model gradients, acting as an "edge preserving" constraint complementing or outweighing data fitting in problematic areas (Anagaw and Sacchi, 2011; Guitton, 2012). However, ℓ_1 and TV regularized optimization problems are difficult to solve, and the development of efficient numerical solution techniques is a subject of active ongoing research, see e.g. Boyd et al. (2011).

In this work we proposed a formulation of the full-waveform inversion as a problem of *constrained optimization*, and solve it using the iterative Bregman regularization technique, see e.g. Osher et al. (2005). We demonstrate advantages of the proposed method over unconstrained regularization. The paper concludes with an example of application to the Marmousi synthetic with added noise.

METHOD

We begin with the standard formulation of FWI as an unconstrained nonlinear least-squares fitting problem (Virieux and Operto, 2009):

$$\|\mathbf{F}(\mathbf{m}) \; - \; \mathbf{d}\|_2 \; \rightarrow \; \min, \tag{1}$$

where $\mathbf{d}$ is the observed data, $\mathbf{m}$ is the model (i.e., acoustic slowness) and $\mathbf{F}$ is the nonlinear forward modeling operator. Problem (1) can be solved in either time or frequency domain, with either approach having its advantages (Virieux and Operto, 2009). Formulation (1) equally weights all data, resulting in better illuminated areas being better constrained. Weighted least-squares and priors may help improve recovery of deeper sections, however, the problem still remains underdetermined at greater depths or acquisition blind spots.

In this work we explore an alternative formulation of FWI as a constrained optimization problem,

$$\begin{aligned} \|\mathbf{Rm}\|_1 \; &\rightarrow \; \min, \\ \mathbf{F}(\mathbf{m}) \; &= \; \mathbf{d}, \end{aligned} \tag{2}$$

where

$$\mathbf{Rm} \; = \; |\nabla_{\mathbf{x}} m|, \tag{3}$$

is the length of spatial slowness gradient, and $\|\mathbf{Rm}\|_1$ is the *total variation* seminorm of $\mathbf{m}$ (Triebel, 2006). Note that in (2) we use an equality constraint for data fitting, which is neither desirable nor realistic in applications to field data. Indeed, formulation (2) is suitable for strictly *underdetermined* problems, and the equality constraint can be enforced only for noise-free data. Since the full-waveform inversion problem is *mixed-determined* and field data are always noisy, we propose to solve the following constrained optimization problem:

$$\begin{aligned} \|w(\mathbf{x})\mathbf{Rm}\|_1 \; &\rightarrow \; \min, \\ \|\mathbf{F}(\mathbf{m}) \; - \; \mathbf{d}\|_2^2 \; &< \; \sigma^2, \end{aligned} \tag{4}$$

where $w(\mathbf{x}) \; = \; w(x^1, x^2, x^3)$ is a weighting function that is above zero only in areas of poor resolution or illumination at depth. In a practical solution algorithm, solving (4) is equivalent to solving (2) with suitable stopping criteria when the desired data misfit σ^2 is achieved.

We solve (4) using the iterative *Bregman regularization* technique proposed by Osher et al. (2005). Starting from

$$\mathbf{m} \; = \; 0, \;\; \mathbf{p}_0 \; = \; 0,$$

and given $\mathbf{m}_k$, we iteratively compute slowness $\mathbf{m}_{k+1}$ as the solution of the following unconstrained TV-regularized problem

$$\lambda\|w(\mathbf{x})\mathbf{Rm}\|_1 \; - \; \langle \mathbf{p}_k, \mathbf{m} - \mathbf{m}_k \rangle + \|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2 \; \rightarrow \; \min, \tag{5}$$

where an element $\mathbf{p}_{k+1}$ of the *subgradient* of $\lambda\|w(\mathbf{x})\mathbf{Rm}\|_1$ is computed as

$$\mathbf{p}_{k+1} \; = \; \mathbf{p}_k \; - \; \nabla_{\mathbf{m}} \|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2 \big|_{\mathbf{m}=\mathbf{m}_{k+1}} . \tag{6}$$

Note that (5) describes a TV-regularized inversion (Rudin et al., 1992) that may yield an edge-preserving or "blocky" approximation to the solution of the full-waveform inversion

problem (1). The first two terms in (5) are known as the *Bregman distance* (Bregman, 1967), that is equivalent to extracting an approximately quadratic function from the regularization term. We use the regularization approach based on solving the single unconstrained problem similar to (5) with $\mathbf{p}_k \equiv 0$ in our work on TV-regularized time-lapse FWI (Maharramov and Biondi, 2014). It is important to note, however, that in this work, solution of an unconstrained TV-regularized problem (5) is just a single iteration of an algorithm for solving the *constrained* problem (4). We solve problem (5) using nonlinear conjugate gradients (NCG) algorithm (Nocedal and Wright, 2006), with the smoothed TV regularization term

$$\|\sqrt{|\nabla_{\mathbf{x}} m|^2 + \epsilon}\|_1, \tag{7}$$

where $\epsilon \approx 10^{-5}$ is chosen as a threshold for realistic values of the slowness.

Choice of the regularization parameter λ in (5) can be based on achieving better conditioning of problem (5), and unlike the traditional penalty function/continuation methods for solving (2), λ does not increase between iterations (Goldstein and Osher, 2009; Cai et al., 2010). Furthermore, iterations (5,6) can be shown to converge to a solution of (2) (or a solution of (4) for some $\sigma > 0$ for noisy data) regardless of the value of $\lambda > 0$, as can be demonstrated by a trivial extension of the technique of Cai et al. (2010). However, the rate of convergence does depend on the value of the regularization parameter λ, making the application of Bregman regularization to some nonlinear operators $\mathbf{F}$ problematic. However, our experiments indicate that the value of the regularization parameter chosen to improve the convergence of NCG for (5) results in good overall convergence of Bregman iterations.

Iterations are stopped when the data misfit reaches a desired value of $\sigma > 0$ (Osher et al., 2005; Cai et al., 2010). In practical applications where σ may not be known a-priori, iterations may continue until the effects of overfitting start exceeding the edge-preserving effects of the regularization term (3). Note that instead of using the NCG to solve (5) with the smoothed regularization term (3), problem (5) can be solved using *split Bregman* method that only requires iterative solution of nonlinear least squares problems and soft thresholding (Goldstein and Osher, 2009). However, our numerical experiments indicate that the NCG applied to the smoothed (3) has equivalent performance and accuracy.

NUMERICAL EXAMPLES

We apply the method to the synthetic dataset used in (Maharramov and Biondi, 2014), generated for the Marmousi velocity model over a 384×122 grid with a 24 m grid spacing. The inversion is carried out in the frequency domain for 3.0, 3.6, 4.3, 5.1, 6.2, 7.5, 9.0, 10.8, 12.8, and 15.5 Hz with time-domain forward modeling (Sirgue et al., 2010). The frequencies are chosen based on the estimated offset to depth range of the data (Sirgue and Pratt, 2004). The acquisition has 192 shots at a depth of 16 m with a 48 m spacing, and 381 receivers at a depth of 15 m with a 24 m spacing. The minimum offset is 48 m. The source function is a Ricker wavelet centered at 10.1 Hz. Absorbing boundary conditions are applied along the entire model boundary, including the surface (thus suppressing multiples). A smoothed true model shown in (Maharramov and Biondi, 2014) is used as a starting model for the inversion. The smoothing is performed using a triangular filter with a 20-sample half-window in both vertical and horizontal directions. Random Gaussian noise is added to the noise-free synthetic data to produce a noisy dataset with 7 dB signal-to-noise ratio. The result of model inversion from the 7 dB SNR synthetic data is shown in Figure 1. Up to 10

iterations of the nonlinear conjugate gradients algorithm (Nocedal and Wright, 2006) are performed for each frequency. Neither regularization nor model priors are used. Figure 2

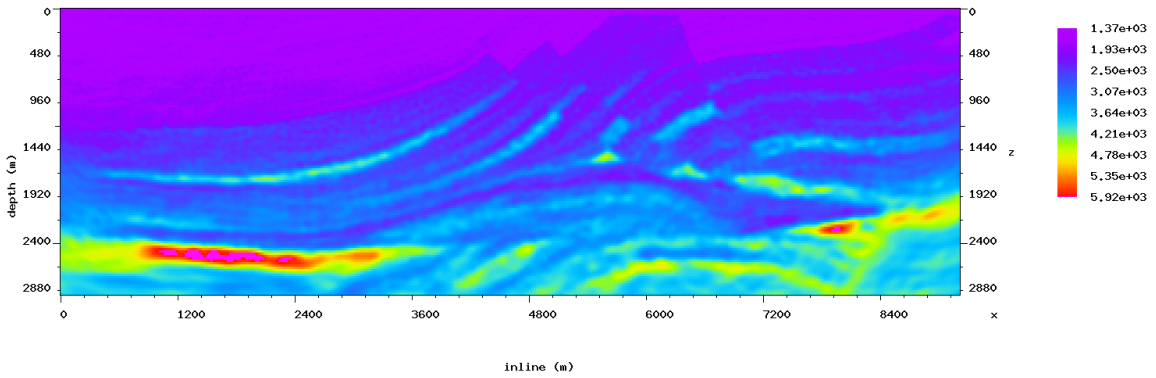


Figure 1: Inversion of 7dB SnR synthetic using the unregularized FWI, 10 iterations per frequency. [**CR**] musa2/. base

shows the results of solving the proposed constrained FWI (4). Only 5 NCG iterations were used for solving each problem (5), with only two outer (Bregman) iterations, resulting in roughly the same compute time as in our standard FWI experiment shown in Figure 1 (10 gradient evaluations using the adjoint state method). The weighting function $w(\mathbf{x})$ was set to 1 below 2100 m and zero above 2000 m, thus the regularization is applied to less constrained areas. Our results in Figure 2 indicate that CFWI has improved the deepest

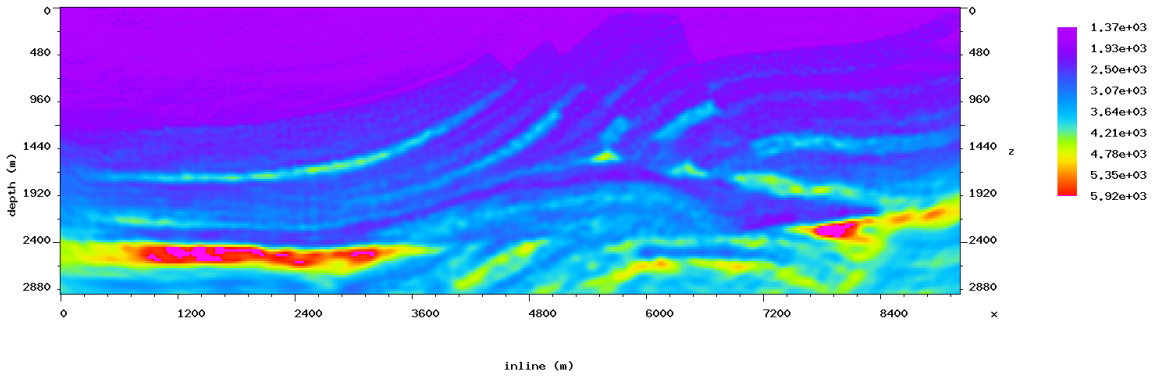


Figure 2: Inversion of 7dB SnR synthetic by solving constrained TV-regularized FWI (4) using Bregman iterative procedure, 10 iterations per frequency. Note that while the shallow parts are similar to Figure 1, deeper sections below 2km are more focused, and the poorly illuminated and mispositioned intervals in the left part of the model have been improved. [**CR**] musa2/. cfwi

section of the model while matching the standard FWI in more shallow well-constrained areas. The result of Figure 2 is closer to the clean synthetic inversion shown in (Maharramov and Biondi, 2014), and has better delineated interfaces.

CONCLUSIONS

We have proposed a new formulation of FWI as a constrained optimization problem (CFWI), and demonstrated the CFWI to be a viable technique for improving depth resolution and accuracy of FWI. Application of Bregman iterative regularization provides a computationally efficient solution method for CFWI that can be easily built on top of the existing solvers. Application of CFWI to field data will be the subject of future work.

ACKNOWLEDGMENTS

The authors would like to thank Jon Claerbout and Stewart Levin for a number of useful discussions, and the Stanford Center for Computational Earth and Environmental Sciences for providing computing resources.

REFERENCES

Anagaw, A. Y. and M. D. Sacchi, 2011, Full waveform inversion with total variation regularization: Presented at the Recovery-CSPG CSEG CWLS Convention.

Boyd, S., N. Parikh, E. Chu, B. Peleato, and J. Eckstein, 2011, Distributed optimization and statistical learning via the alternating direction method of multipliers: Foundations and Trends in Machine Learning, **3**, 1–122.

Bregman, L., 1967, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming: {USSR} Computational Mathematics and Mathematical Physics, **7**, 200 – 217.

Cai, J., S. Osher, and Z. Shen, 2010, Split bregman methods and frame based image restoration: Multiscale Modeling & Simulation, **8**, 337–369.

Goldstein, T. and S. Osher, 2009, The split bregman method for l1-regularized problems: SIAM Journal on Imaging Sciences, **2**, 323–343.

Guitton, A., 2012, Blocky regularization schemes for full-waveform inversion: Geophysical Prospecting, **60**, 870–884.

Maharramov, M. and B. Biondi, 2014, Robust joint full-waveform inversion of time-lapse seismic datasets with total-variation regularization: SEP Report, **155**, 199–208.

Nocedal, J. and S. J. Wright, 2006, Numerical optimization: Springer.

Osher, S., M. Burger, D. Goldfarb, J. Xu, and W. Yin, 2005, An iterative regularization method for total variation-based image restoration: Multiscale Modeling & Simulation, **4**, 460–489.

Rudin, L. I., S. Osher, and E. Fatemi, 1992, Nonlinear total variation based noise removal algorithms: Phys. D, **60**, 259–268.

Sirgue, L., O. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. Kommedal, 2010, Full waveform inversion: the next leap forward in imaging at Valhall: First Break, **28**, 65–70.

Sirgue, L. and R. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248.

Triebel, H., 2006, Theory of function spaces: Birkh auser.

Virieux, J. and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26.

Tutorial on two-way wave equation operators for acoustic, isotropic, constant-density media

Guillaume Barnier and Ali Almomin

ABSTRACT

This paper is a tutorial on linearized two-way wave equation modeling and inversion operators. We provide a detailed derivation for the special case of an acoustic, isotropic, constant-density medium. We analyze the Born, tomographic, and WEMVA forward modeling operators, their adjoints, and we extend the analysis to the subsurface offset domain.

INTRODUCTION

We present a detailed derivation for the main two-way wave equation linearized operators used in seismic imaging: Born, tomography, and Wave Equation Migration Velocity Analysis (WEMVA). This work adapts that of Almomin (2013) and is intended to serve as an educational tool for new students coming to SEP. We will consider the specific case of an acoustic, isotropic, constant-density medium.

Our goal is to obtain the best possible estimate of the seismic velocity model of the Earth's subsurface. The three operators that we derive are convenient to achieve this goal. The Born operator and its adjoint capture the dynamic effects responsible for seismic reflections (the high wavenumber content of the velocity model). The tomographic and WEMVA operators and their adjoints capture the kinematic effects (e.g., transmitted and diving waves), which are controlled by the low wavenumber content of the velocity model.

The first section is intended to remind the reader of some general background on wave theory. We then derive the Born modeling operator and its adjoint, referred to as the Reverse Time Migration (RTM) operator. In the last two sections, we will treat the tomographic and WEMVA operators and their respective adjoints.

Throughout this paper, we use roman fonts to refer to functions, italic fonts to refer to functions evaluated at a given point, and bold fonts to refer to vectors (e.g., f, $f(x)$, and $\mathbf{f}$). Moreover, we use lowercase to refer to functions in the time domain, and uppercase for their Discrete Fourier Transform (DFT).

WAVE THEORY

Definitions

- m is a function representing the seismic velocity model of the subsurface. For each point in the subsurface, it associates a velocity value:

$$\begin{aligned} \mathrm{m} : \Omega &\mapsto \mathbb{R} \\ \mathbf{x} &\mapsto m(\mathbf{x}), \end{aligned} \tag{1}$$

where $m(\mathbf{x})$ is expressed in m/s. $\Omega \subset \mathbb{R}^3$ is the area of study, and is an open, bounded and regular set. It is assumed that m does not vary with time.

- p is the pressure field function. For a given location in the subsurface $\mathbf{x}$, at a given time t, and for a given velocity function m, it represents the pressure value

$$\begin{aligned} \mathrm{p} : \Omega \times \mathbb{R} \times \mathcal{F}(\Omega, \mathbb{R}) &\mapsto \mathbb{R} \\ (\mathbf{x}, t; \mathrm{m}) &\mapsto p(\mathbf{x}, t; \mathrm{m}), \end{aligned} \tag{2}$$

where $p(\mathbf{x}, t; \mathrm{m})$ is expressed in Pa, and $\mathcal{F}(\Omega, \mathbb{R})$ is the set of functions mapping Ω to $\mathbb{R}$ (assumed to be infinitely differentiable on Ω).

- s is a function representing a seismic source:

$$\begin{aligned} \mathrm{s} : \Omega \times \mathbb{R} &\mapsto \mathbb{R} \\ (\mathbf{x}, t) &\mapsto s(\mathbf{x}, t), \end{aligned} \tag{3}$$

where $s(\mathbf{x}, t)$ is expressed in $\mathrm{Pa/m^2}$.

- ρ is a function representing the medium volume mass density and is defined by

$$\begin{aligned} \rho : \Omega &\mapsto \mathbb{R} \\ \mathbf{x} &\mapsto \rho(\mathbf{x}), \end{aligned} \tag{4}$$

where $\rho(\mathbf{x})$ is expressed in $\mathrm{kg/m^3}$. In the following, we assume ρ to be constant over the study area.

- $\forall \mathbf{x} \in \Omega, \forall t \in \mathbb{R}$, the acoustic, isotropic constant-density wave equation satisfied by the causal function p is given by

$$\begin{aligned} \frac{1}{m^2(\mathbf{x})} \frac{\partial^2 p(\mathbf{x}, t; \mathrm{m})}{\partial t^2} - \nabla^2 p(\mathbf{x}, t; \mathrm{m}) &= s(\mathbf{x}, t) \\ p(\mathbf{x}, t; \mathrm{m}) &\equiv 0, \ \forall t \leq 0. \end{aligned} \tag{5}$$

Continuous solution of the wave equation

Green's function

Given a source located at $\mathbf{x}'$ and initiated at time t', the causal Green's function, g, for an acoustic, isotropic constant-density medium is the solution to equation 5, where s is a point source function, and an impulse in time. That is, $\forall \mathbf{x} \in \Omega,\ \forall t \in \mathbb{R}, \forall t' \in \mathbb{R}$, g satisfies

$$\begin{aligned} \frac{1}{m^2(\mathbf{x})} \frac{\partial^2 g(\mathbf{x}, t, \mathbf{x}', t'; \mathrm{m})}{\partial t^2} - \nabla^2 g(\mathbf{x}, t, \mathbf{x}', t'; \mathrm{m}) &= \delta(\mathbf{x} - \mathbf{x}')\delta(t - t') \\ g(\mathbf{x}, t, \mathbf{x}', t'; \mathrm{m}) &\equiv 0,\ \forall t \leq t', \end{aligned} \tag{6}$$

where

- $\delta(\mathbf{x} - \mathbf{x}')$ is in m^{-3} (for 3D),
- $\delta(t - t')$ is in s^{-1}, and
- $g(\mathbf{x}, t, \mathbf{x}', t'; \mathrm{m})$ is in $\mathrm{m}^{-1}\,\mathrm{s}^{-1}$ (and not Pa).

Spatial reciprocity implies that

$$g(\mathbf{x}, t, \mathbf{x}', t'; \mathrm{m}) = g(\mathbf{x}', t, \mathbf{x}, t'; \mathrm{m}). \tag{7}$$

Assuming the properties of the medium are invariant with time, we can write

$$g(\mathbf{x}, t, \mathbf{x}', t'; \mathrm{m}) = \mathrm{g}(\mathbf{x}, t - t', \mathbf{x}', 0; \mathrm{m}). \tag{8}$$

$\forall \mathbf{x} \in \Omega, \forall t \in \mathbb{R}$, the solution of equation 5 can be written as

$$p(\mathbf{x}, t; \mathrm{m}) = \int\limits_{\mathbf{x}' \in \Omega} \int\limits_{t'=-\infty}^{+\infty} g(\mathbf{x}, t, \mathbf{x}', t'; \mathrm{m})\ s(\mathbf{x}', t')\ \mathrm{d}t'\ \mathrm{d}\mathbf{x}'. \tag{9}$$

Using equation 8, we may write

$$\int\limits_{t'=-\infty}^{+\infty} g(\mathbf{x}, t, \mathbf{x}', t'; \mathrm{m})\ s(\mathbf{x}', t')\ \mathrm{d}t' = g(\mathbf{x}, t, \mathbf{x}', 0; \mathrm{m}) * s(\mathbf{x}', t), \tag{10}$$

where * denotes convolution in time. Equation 9 thereby simplifies to

$$p(\mathbf{x},t;\mathrm{m}) = \int\limits_{\mathbf{x}'\in\Omega} g(\mathbf{x},t,\mathbf{x}',0;\mathrm{m}) * s(\mathbf{x}',t)\ \mathrm{d}\mathbf{x}'. \tag{11}$$

If the seismic source function s is concentrated at one point in space $\mathbf{x}_s$, and has a time signature $f(t)$, then $s(\mathbf{x},t) = f(t)\ \delta(\mathbf{x}-\mathbf{x}_s)$, where $f(t)$ is expressed in Pa m, and $\delta(\mathbf{x}-\mathbf{x}_s)$ is expressed in m^{-3}. In that case, equation 11 simplifies further to

$$p(\mathbf{x},t;\mathrm{m}) = g(\mathbf{x},t,\mathbf{x}_s;\mathrm{m}) * f(t),\ \forall\mathbf{x}\in\Omega, \forall t\in\mathbb{R}. \tag{12}$$

Numerical solution of the wave equation

Discretization in time and frequency

We discretize all functions/signals in time. Hence $\forall n \in \{1;N\}$,

$$p(\mathbf{x},n;\mathrm{m}) \quad = \quad p(\mathbf{x},t_n;\mathrm{m}), \tag{13}$$

where dt is the time sampling rate in s/samples, and $t_n = (n-1)dt$. Here, it is assumed that $\exists N\in\mathbb{N}$ such that $n\notin\{1;N\} \Rightarrow p(\mathbf{x},n;\mathrm{m}) = 0$.

By taking the Discrete Fourier Transform (DFT) of each side of equation 11, we obtain that $\forall k\in\{1;N\}$,

$$P(\mathbf{x},\omega_k;\mathrm{m}) = \int\limits_{\mathbf{x}'\in\Omega} G(\mathbf{x},\omega_k,\mathbf{x}';\mathrm{m})\ S(\mathbf{x}',\omega_k)\ \mathrm{d}\mathbf{x}', \tag{14}$$

where $\omega_k = (k-1)d\omega = (k-1)\dfrac{2\pi}{N}$. Note that $P(\mathbf{x},\omega_k;\mathrm{m})\in\mathbb{C}$. Moreover,

- $P(\mathbf{x},\omega_k;\mathrm{m})$ is expressed in Pa s,
- $F(\omega_k)$ is expressed in Pa m s, and
- $G(\mathbf{x},\omega_k,\mathbf{x}_s;\mathrm{m})$ is expressed in m^{-1}.

Discretization in space

We discretize the area of study into a regularly sampled grid. $\forall n\in\{1;N\},\ \forall j\in\{1;M\}$, equation 11 becomes

$$p(\mathbf{x}, n; \mathbf{m}) \approx \sum_{j=1}^{M} g(\mathbf{x}, n, \mathbf{x}_j; \mathrm{m}) * s(\mathbf{x}_j, n)\ \Delta\mathbf{x}, \tag{15}$$

where $\Delta\mathbf{x}$ is the constant finite difference grid cell volume, expressed in m^3. Since $\Delta\mathbf{x}$ acts as a constant scaling factor, we will not write it explicitly in the remaining of our derivation. However, this coefficient is important to ensure consistency in the units. The bold font used for the model function $\mathbf{m}$ indicates that we have spatially discretized the study area. M is the number of grid points in the study area (discretized model size). Therefore, we can define a model vector $\mathbf{m} \in \mathbb{R}^M$, whose components are the values of function m (defined in mapping 1), evaluated at each grid point $\mathbf{x}_i$

$$\mathbf{m} = \begin{pmatrix} m(\mathbf{x}_1) \\ \vdots \\ m(\mathbf{x}_M) \end{pmatrix} = \begin{pmatrix} m_1 \\ \vdots \\ m_M \end{pmatrix}. \tag{16}$$

We can also express equation 15 in the frequency domain. $\forall i \in \{1; M\},\ \forall k \in \{1; N\}$,

$$P(\mathbf{x}_i, \omega_k; \mathbf{m}) \ \approx \ \sum_{j=1}^{M} G(\mathbf{x}_i, \omega_k, \mathbf{x}_j; \mathbf{m})\ S(\mathbf{x}_j, \omega_k). \tag{17}$$

If the seismic acquisition source is concentrated at a point $\mathbf{x}_s$ in space, its DFT is expressed by $S(\mathbf{x}_i, \omega_k) = F(\omega_k)\ \delta(\mathbf{x}_i - \mathbf{x}_s)$, where F is the DFT of the source's time signature. In that case, equation 17 simplifies further to

$$P(\mathbf{x}_i, \omega_k; \mathbf{m}) = G(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{m})\ F(\omega_k). \tag{18}$$

Seismic data

Seismic data are a set of discrete measurements in time and space *received* at some locations $\mathbf{x}_i$ (e.g., near the surface of the Earth). For one seismic source located at $\mathbf{x}_s$, we defined a data vector $\mathbf{D} \in \mathbb{C}^{N_d \times N}$ (N_d is the number of receiver locations)

$$\mathbf{D} = \begin{pmatrix} D_{11} \\ \vdots \\ D_{N_d N} \end{pmatrix}. \tag{19}$$

Each component D_{ik} of $\mathbf{D}$ can be expressed by

$$D_{ik} = P(\mathbf{x} = \mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{m}), \tag{20}$$

where $i \in \{1; N_d\},\ k \in \{1; N\}$.

Wave equation and Green's function in the frequency domain

By discretizing in space and taking the DFT of equation 6, the Green's function G satisfies the Helmholtz equation. $\forall i \in \{1; M\}, \forall k \in \{1; N\}$,

$$\left[\frac{\omega_k^2}{m(\mathbf{x}_i)^2} + \nabla^2\right] G(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{m}) = -\delta(\mathbf{x}_i - \mathbf{x}_s). \tag{21}$$

The units of equation 21 are consistent, as each side is expressed in m^{-3}.

LINEARIZATION WITH RESPECT TO REFLECTIVITY

Nonlinear mapping

One way to think about the wave equation is as a function f that maps the set of model parameters (velocity value at each grid point) to the recorded seismic data

$$\begin{aligned} \mathrm{f} : \mathbb{R}^M &\mapsto \mathbb{R}^{N_d} \\ \mathbf{m} &\mapsto \mathbf{f}(\mathbf{m}), \end{aligned} \tag{22}$$

where $\mathbf{d} = \mathbf{f}(\mathbf{m})$ is the seismic data in the time domain. Equivalently, in the frequency domain,

$$\begin{aligned} \mathrm{F} : \mathbb{R}^M &\mapsto \mathbb{C}^{N_d} \\ \mathbf{m} &\mapsto \mathbf{F}(\mathbf{m}), \end{aligned} \tag{23}$$

where $\mathbf{D} = \mathbf{F}(\mathbf{m})$ is the seismic data in the frequency domain. Clearly, f and F (to be distinguished from the source signatures mentioned previously) are not linear functions with respect to $\mathbf{m}$. However, they are both linear operators with respect to the *source function*, keeping the velocity model and all other variables unchanged. Let us represent the wave equation operator by $\mathcal{L}$ such that

$$\begin{aligned} \mathcal{L} : \mathcal{F}(\Omega, \mathbb{R}) &\mapsto \mathcal{F}(\Omega, \mathbb{R}) \\ \mathrm{s} &\mapsto \mathcal{L}(\mathrm{s}). \end{aligned} \tag{24}$$

Here, $\mathrm{s} \in \mathcal{F}(\Omega, \mathbb{R})$ is a source function. One can easily verify from equation 5 that, $\forall \alpha \in \mathbb{R}, \forall\ (\mathrm{s}_i, \mathrm{s}_j) \in (\mathcal{F}(\Omega, \mathbb{R}))^2$

$$\begin{aligned} \mathcal{L}(\mathrm{s}_i + \mathrm{s}_j) &= \mathcal{L}(\mathrm{s}_i) + \mathcal{L}(\mathrm{s}_j) \\ \mathcal{L}(\alpha\, \mathrm{s}_i) &= \alpha \mathcal{L}(\mathrm{s}_i). \end{aligned} \tag{25}$$

The next step is to linearize F with respect to $\mathbf{m}$. In the following, we present our derivation in the frequency domain. First, we assume that we can decompose the model vector $\mathbf{m} \in \mathbb{R}^M$ as the sum of two vectors

$$\mathbf{m} = \mathbf{b} + \mathbf{r}, \tag{26}$$

where

$$\mathbf{b} = \begin{pmatrix} b(\mathbf{x}_1) \\ \vdots \\ b(\mathbf{x}_M) \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_M \end{pmatrix}, \tag{27}$$

and

$$\mathbf{r} = \begin{pmatrix} r(\mathbf{x}_1) \\ \vdots \\ r(\mathbf{x}_M) \end{pmatrix} = \begin{pmatrix} r_1 \\ \vdots \\ r_M \end{pmatrix}. \tag{28}$$

$\mathbf{b} \in \mathbb{R}^M$ is referred to as the *background* model, and contains the low wavenumber content of the velocity model. $\mathbf{r} \in \mathbb{R}^M$ is referred to as the *reflectivity* model, and contains the high wavenumber content of the velocity model. We also assume that the magnitude of the reflectivity is much smaller than the one of the background, $\|\mathbf{r}\| \ll \|\mathbf{b}\|$, where $\|.\|$ denotes any norm on $\mathbb{R}^M$. In the following, we will treat those parameters as two separate variables with same units ($\mathrm{m\,s^{-1}}$).

Let us consider the function

$$\begin{aligned} \tilde{\mathrm{F}} : \mathbb{R}^M &\mapsto \mathbb{C}^{N_d} \\ \mathbf{r} &\mapsto \tilde{\mathbf{F}}(\mathbf{r}), \end{aligned} \tag{29}$$

which is the restriction of F to the high wavenumber part of the velocity model, while keeping the background velocity model unchanged. We perform a multivariate Taylor expansion of $\tilde{\mathrm{F}}$ around a reflectivity model $\mathbf{r}_0$ (referred to as the background reflectivity) while keeping the background model $\mathbf{b}$ unchanged

$$\tilde{\mathbf{F}}(\mathbf{r}) = \tilde{\mathbf{F}}(\mathbf{r}_0) + \left.\frac{\partial \tilde{\mathbf{F}}(\mathbf{r})}{\partial \mathbf{r}}\right|_{\mathbf{r}=\mathbf{r}_0} \Delta\mathbf{r} + \mathcal{O}(\|\Delta\mathbf{r}\|^2). \tag{30}$$

The second term of the right side of equation 30 is the Jacobian matrix $\mathcal{B}$ of $\tilde{\mathbf{F}}$ evaluated at $\mathbf{r} = \mathbf{r}_0$, applied to the reflectivity perturbation vector $\Delta\mathbf{r}$. It can be expressed by

$$\mathcal{B}(\mathbf{r}_0) = \frac{\partial \tilde{\mathbf{F}}(\mathbf{r})}{\partial \mathbf{r}}\bigg|_{\mathbf{r}=\mathbf{r}_0} = \begin{pmatrix} \frac{\partial \tilde{F}_1}{\partial r_1}\big|_{\mathbf{r}_0} & \cdots & \frac{\partial \tilde{F}_1}{\partial r_M}\big|_{\mathbf{r}_0} \\ \vdots & \ddots & \vdots \\ \frac{\partial \tilde{F}_{N_d}}{\partial r_1}\big|_{\mathbf{r}_0} & \cdots & \frac{\partial \tilde{F}_{N_d}}{\partial r_M}\big|_{\mathbf{r}_0} \end{pmatrix}, \tag{31}$$

where $\mathcal{B}(\mathbf{r}_0) \in M_{N_d,M}(\mathbb{C})$ ($M_{p,q}(\mathbb{C})$ refers to the set of matrices with p rows, q columns, and complex coefficients).

Born operator

The Born operator is a linear operator that relates a perturbation in the model reflectivity to a perturbation in the data (Almomin, 2013), while keeping the background unchanged.

Expressing the entries of the Jacobian

We consider a reflectivity perturbation $\Delta\mathbf{r}$ vector such that

- $\mathbf{r} = \mathbf{r}_0 + \Delta\mathbf{r}$, and
- $\|\Delta\mathbf{r}\| \ll \|\mathbf{r}_0\|$.

Thus,

$$\begin{aligned} \Delta\mathbf{D} &= \mathbf{D}(\mathbf{r}) - \mathbf{D}(\mathbf{r}_0) \\ &\approx \mathcal{B}(\mathbf{r}_0)\,\Delta\mathbf{r}. \end{aligned} \tag{32}$$

The Born operator is a linear application $\mathcal{B}(\mathbf{r}_0) : \mathbb{C}^M \mapsto \mathbb{C}^{N_d}$, which can be represented in a matrix form by $\mathcal{B}(\mathbf{r}_0)$. It is the Jacobian of $\tilde{\mathbf{F}}$ taken at $\mathbf{r}_0$, whose entries are all independent of the perturbation $\Delta\mathbf{r}$.

Let us define, $\forall i \in \{1; N_d\}, \forall k \in \{1; N\}$,

$$D_{ik}(\mathbf{b}, \mathbf{r}) = P(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r}) = F(\omega_k)\, G(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r}), \tag{33}$$

where $D_{ik}(\mathbf{b}, \mathbf{r}) \in \mathbb{C}$. Therefore, we can rewrite equation 32 with the following form:

$$\begin{aligned} \Delta D(\mathbf{x}_i, \mathbf{x}_s, \omega_k, ; \mathbf{b}, \mathbf{r}) &= P(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r}) - P(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r}_0) \\ &\approx \sum_{j=1}^{M} \frac{\partial P}{\partial r_j}(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r})\bigg|_{\mathbf{r}=\mathbf{r}_0} \Delta r(\mathbf{x}_j) \\ &\approx F(\omega_k) \sum_{j=1}^{M} \frac{\partial G}{\partial r_j}(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r})\bigg|_{\mathbf{r}=\mathbf{r}_0} \Delta r(\mathbf{x}_j). \end{aligned} \tag{34}$$

The left side of equation 34 is the data perturbation caused by the perturbation of the reflectivity model, for one fixed frequency, one fixed receiver location, one fixed source, and a fixed background velocity model.

To find an expression of $\frac{\partial G}{\partial r_j}$, we write two versions of equation 21: one with the unperturbed velocity model $\mathbf{m}_0 = \mathbf{b} + \mathbf{r}_0$, satisfied by $G(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r}_0)$ (equation 35), and the other with the perturbed velocity model $\mathbf{m} = \mathbf{m}_0 + \Delta\mathbf{r}$, satisfied by $G(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r})$. That gives us, $\forall i \in \{1; M\}, \forall k \in \{1; N\}$,

$$\left[\frac{\omega_k^2}{\big(m_0(\mathbf{x}_i) + \Delta r(\mathbf{x}_i)\big)^2} + \nabla^2\right] G(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r}) \quad = \quad -\delta(\mathbf{x}_i - \mathbf{x}_s), \tag{35}$$

and

$$\left[\frac{\omega_k^2}{m_0(\mathbf{x}_i)^2} + \nabla^2\right] G(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r}_0) \quad = \quad -\delta(\mathbf{x}_i - \mathbf{x}_s). \tag{36}$$

Since $\forall i \in \{1; M\}$, $|\Delta r(\mathbf{x}_i)| \ll |m_0(\mathbf{x}_i)|$, we can make the following approximation:

$$\frac{1}{\big(m_0(\mathbf{x}_i) + \Delta r(\mathbf{x}_i)\big)^2} \approx \frac{1}{m_0(\mathbf{x}_i)^2}\left(1 - \frac{2\Delta r(\mathbf{x}_i)}{m_0(\mathbf{x}_i)}\right). \tag{37}$$

By expanding equations 35, 36, and 37, we can show that the difference $\Delta G(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r}) = G(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r}) - G(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r}_0)$ satisfies a similar equation as in equation 21, but for a different source function. $\forall i \in \{1; M\}, \forall k \in \{1; N\}$,

$$\begin{aligned}\left[\frac{\omega_k^2}{m_0(\mathbf{x}_i)^2} + \nabla^2\right] \Delta G(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r}) = \\ 2\,\omega_k^2\,\frac{\Delta r(\mathbf{x}_i)}{m_0(\mathbf{x}_i)^3}\,\Big(G(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r}_0) + \Delta G(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r})\Big).\end{aligned} \tag{38}$$

Under that form, this wave equation is not linear with respect to the source (mapping 24). However, we can further simplify it by noticing in equation 38 that the term proportional to $\Delta r(\mathbf{x}_i)\Delta G$ is a second order differential element, and can thus be neglected with respect to the remaining terms (Born approximation). Hence,

$$|\Delta r(\mathbf{x}_i)\Delta G| \ll |\Delta r(\mathbf{x}_i)\, G_0|, \tag{39}$$

and

$$\left|\frac{2\,\omega_k^2}{m_0(\mathbf{x}_i)^3}\Delta r(\mathbf{x}_i)\Delta G\right| \ll \left|\Big[\frac{\omega_k^2}{m_0(\mathbf{x}_i)^2} + \nabla^2\Big]\Delta G\right|. \tag{40}$$

Therefore, equation 38 simplifies to

$$\left[\frac{\omega_k^2}{m_0(\mathbf{x}_i)^2} + \nabla^2\right] \Delta G \approx S_{\text{sec}}(\mathbf{x}_i, \omega_k), \tag{41}$$

with

$$S_{\text{sec}}(\mathbf{x}_i, \omega_k) = 2\,\omega_k^2\,\frac{\Delta r(\mathbf{x}_i)}{m_0(\mathbf{x}_i)^3}\; G(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r}_0). \tag{42}$$

ΔG satisfies the wave equation with a secondary source function S_{sec}, *independent* of ΔG, and proportional to the reflectivity perturbation at a given point in the subsurface. Using equation 17, we can solve for ΔG. $\forall i \in \{1; M\}, \forall k \in \{1; N\}$,

$$\begin{aligned} \Delta G(\mathbf{x}_i, \omega_k; \mathbf{b}, \mathbf{r}) &= \sum_{j=1}^{M} G(\mathbf{x}_i, \omega_k, \mathbf{x}_j; \mathbf{b}, \mathbf{r}_0)\; S_{\text{sec}}(\mathbf{x}_j, \omega_k) \\ &= \sum_{j=1}^{M} \frac{2\,\omega_k^2}{m_0(\mathbf{x}_j)^3} G(\mathbf{x}_i, \omega_k, \mathbf{x}_j; \mathbf{b}, \mathbf{r}_0)\, G(\mathbf{x}_j, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r}_0)\; \Delta r(\mathbf{x}_j). \end{aligned} \tag{43}$$

We check for the consistency of the units in equation 43, keeping in mind that we did not explicitly write the grid cell volume term (expressed in m^3) on the right side,

- $\Delta\, G(\mathbf{x}_i, \omega_k; \mathbf{b}, \mathbf{r})$ is expressed in m^{-1},
- $G(\mathbf{x}_i, \omega_k, \mathbf{x}_j; \mathbf{b}, \mathbf{r}_0)$ is in m^{-1}, and
- $S_{\text{sec}}(\mathbf{x}_i, \omega_k)$ is in m^{-3}.

The Taylor expansion of ΔG, expressed by

$$\begin{aligned} \Delta G(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r}) &= G(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r}) - G(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r}_0) \\ &\approx \sum_{j=1}^{M} \frac{\partial G}{\partial r_j}(\mathbf{x}_i, \omega_k; \mathbf{x}_s; \mathbf{b}, \mathbf{r})\bigg|_{\mathbf{r}=\mathbf{r}_0} \Delta r(\mathbf{x}_j), \end{aligned} \tag{44}$$

enables us to identify $\dfrac{\partial G}{\partial r_j}$, where

$$\frac{\partial G}{\partial r_j}(\mathbf{x}_i, \omega_k; \mathbf{x}_s; \mathbf{b}, \mathbf{r})\bigg|_{\mathbf{r}=\mathbf{r}_0} = \frac{2\,\omega_k^2}{m_0(\mathbf{x}_j)^3} G(\mathbf{x}_i, \omega_k, \mathbf{x}_j; \mathbf{b}, \mathbf{r}_0)\, G(\mathbf{x}_j, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r}_0). \tag{45}$$

Finally, we can express the perturbation of the data with respect to the perturbation of the reflectivity. $\forall i \in \{1; N_d\}, \forall k \in \{1; N\}$,

$$\begin{aligned} \Delta D_{ik} &= \Delta D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r}) \\ &= \sum_{j=1}^{M} F(\omega_k) \frac{2\,\omega_k^2}{m_0(\mathbf{x}_j)^3} G(\mathbf{x}_i, \omega_k, \mathbf{x}_j; \mathbf{b}, \mathbf{r}_0) \Delta r(\mathbf{x}_j)\, G(\mathbf{x}_j, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r}_0). \end{aligned} \tag{46}$$

Equation 46 can be written in a matrix form

$$\Delta \mathbf{D} = \mathcal{B}(\mathbf{r}_0) \Delta \mathbf{r}, \tag{47}$$

and

$$\Delta \mathbf{D} = \begin{pmatrix} \Delta D(\mathbf{x}_1, \mathbf{x}_s, \omega_1; \mathbf{b}, \mathbf{r}) \\ \Delta D(\mathbf{x}_2, \mathbf{x}_s, \omega_1; \mathbf{b}, \mathbf{r}) \\ \vdots \\ \Delta D(\mathbf{x}_{N_d}, \mathbf{x}_s, \omega_1; \mathbf{b}, \mathbf{r}) \\ \vdots \\ \Delta D(\mathbf{x}_1, \mathbf{x}_s, \omega_N; \mathbf{b}, \mathbf{r}) \\ \vdots \\ \Delta D(\mathbf{x}_{N_d}, \mathbf{x}_s, \omega_N; \mathbf{b}, \mathbf{r}) \end{pmatrix} = \begin{pmatrix} b_{11}^1 & \cdots & b_{11}^M \\ b_{21}^1 & \cdots & b_{21}^M \\ \vdots & \vdots & \vdots \\ b_{N_d1}^1 & \cdots & b_{N_d1}^M \\ \vdots & \vdots & \vdots \\ b_{1N}^1 & \cdots & b_{1N}^M \\ \vdots & \vdots & \vdots \\ b_{N_dN}^1 & \cdots & b_{N_dN}^M \end{pmatrix} \begin{pmatrix} \Delta r(\mathbf{x}_1) \\ \Delta r(\mathbf{x}_2) \\ \vdots \\ \Delta r(\mathbf{x}_M) \end{pmatrix}, \tag{48}$$

where $\mathcal{B}(\mathbf{r}_0)$ is the Born modeling operator, linearized around $\mathbf{m}_0 = \mathbf{b} + \mathbf{r}_0$. Moreover,

- $\Delta \mathbf{D} \in \mathbb{C}^{N_d \times N}$,
- $\mathcal{B}(\mathbf{r}_0) \in M_{N_d,M}(\mathbb{C})$,
- $b\,{}_{ik}^{j} \in \mathbb{C}$, and
- $\Delta \mathbf{r} \in \mathbb{R}^M$.

Each entry $b\,{}_{ik}^{j}$ of $\mathcal{B}(\mathbf{r}_0)$ is independent of the reflectivity perturbation $\Delta \mathbf{r}$, and can be expressed $\forall i \in \{1; N_d\}, \forall k \in \{1; N\}, \forall j \in \{1; M\}$ by

$$b\,{}_{ik}^{j} = F(\omega_k) \frac{2\,\omega_k^2}{m_0(\mathbf{x}_j)^3} G(\mathbf{x}_i, \omega_k, \mathbf{x}_j; \mathbf{b}, \mathbf{r}_0)\, G(\mathbf{x}_j, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r}_0). \tag{49}$$

Extension to subsurface offset domain

We extend the reflectivity part of the model to the subsurface offset domain (Almomin, 2013), and define an extended reflectivity function $\tilde{\mathrm{r}}$ by

$$\begin{aligned} \tilde{\mathrm{r}} : \Omega \times \mathcal{H} &\mapsto \mathbb{R} \\ \mathbf{x} \times \mathbf{h} &\mapsto \tilde{r}(\mathbf{x}, \mathbf{h}), \end{aligned} \tag{50}$$

where both Ω and $\mathcal{H}$ are subsets of $\mathbb{R}^3$. We discretize the domain on which $\tilde{\mathrm{r}}$ is defined as in equation 28, and we obtain the reflectivity vector $\tilde{\mathbf{r}} \in \mathbb{R}^{M \times N_h}$ extended to the subsurface offset domain, where $N_h = 2\,h + 1$. Moreover, $\forall p \in \{-h; h\}$, $\mathbf{h}_{\text{-}p} = -\mathbf{h}_p$, and

$$\tilde{\mathbf{r}} = \begin{pmatrix} \tilde{r}(\mathbf{x}_1, \mathbf{h}_{\text{-}h}) \\ \vdots \\ \tilde{r}(\mathbf{x}_M, \mathbf{h}_{\text{-}h}) \\ \tilde{r}(\mathbf{x}_1, \mathbf{h}_{\text{-}h+1}) \\ \vdots \\ \tilde{r}(\mathbf{x}_M, \mathbf{h}_h) \end{pmatrix} = \begin{pmatrix} \tilde{r}(\mathbf{x}_1, -\mathbf{h}_h) \\ \vdots \\ \tilde{r}(\mathbf{x}_M, -\mathbf{h}_h) \\ \tilde{r}(\mathbf{x}_1, -\mathbf{h}_{h\text{-}1}) \\ \vdots \\ \tilde{r}(\mathbf{x}_M, \mathbf{h}_h) \end{pmatrix} \tag{51}$$

Equation 46 can then be modified to take into account the subsurface offset dimension (Almomin, 2013). $\forall i \in \{1; N_d\}, \forall k \in \{1; N\}$, we have

$$\begin{aligned} \Delta D_{ik} &= \Delta D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \tilde{\mathbf{r}}) \\ &= \sum_{j=1}^{M} \sum_{p=\text{-}h}^{h} F(\omega_k) \frac{2\,\omega_k^2}{m_0(\mathbf{x}_j)^3}\, G(\mathbf{x}_i, \omega_k, \mathbf{x}_j + \mathbf{h}_p; \mathbf{b}, \tilde{\mathbf{r}}_0) \Delta\tilde{r}(\mathbf{x}_j, \mathbf{h}_p)\, G(\mathbf{x}_j - \mathbf{h}_p, \omega_k, \mathbf{x}_s; \mathbf{b}, \tilde{\mathbf{r}}_0). \end{aligned} \tag{52}$$

We can display equation 52 in the matrix form

$$\Delta\mathbf{D} = \tilde{\mathcal{B}}(\tilde{\mathbf{r}}_0)\ \Delta\tilde{\mathbf{r}}, \tag{53}$$

and

$$\Delta\mathbf{D} = \begin{pmatrix} \Delta D(\mathbf{x}_1,\mathbf{x}_s,\omega_1;\mathbf{b},\tilde{\mathbf{r}}) \\ \Delta D(\mathbf{x}_2,\mathbf{x}_s,\omega_1;\mathbf{b},\tilde{\mathbf{r}}) \\ \vdots \\ \Delta D(\mathbf{x}_{N_d},\mathbf{x}_s,\omega_1;\mathbf{b},\tilde{\mathbf{r}}) \\ \Delta D(\mathbf{x}_1,\mathbf{x}_s,\omega_2;\mathbf{b},\tilde{\mathbf{r}}) \\ \vdots \\ \Delta D(\mathbf{x}_1,\mathbf{x}_s,\omega_N;\mathbf{b},\tilde{\mathbf{r}}) \\ \vdots \\ \Delta D(\mathbf{x}_{N_d},\mathbf{x}_s,\omega_N;\mathbf{b},\tilde{\mathbf{r}}) \end{pmatrix} = \begin{pmatrix} \tilde{b}_{11}^{1(-h)} & \cdots & \tilde{b}_{11}^{M(-h)} & \tilde{b}_{11}^{1(-h+1)} & \cdots & \tilde{b}_{11}^{Mh} \\ \tilde{b}_{21}^{1(-h)} & \cdots & \tilde{b}_{21}^{M(-h)} & \tilde{b}_{21}^{1(-h+1)} & \cdots & \tilde{b}_{21}^{Mh} \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ \tilde{b}_{N_d1}^{1(-h)} & \cdots & \tilde{b}_{N_d1}^{M(-h)} & \tilde{b}_{N_d1}^{1(-h+1)} & \cdots & \tilde{b}_{N_d1}^{Mh} \\ \tilde{b}_{12}^{1(-h)} & \cdots & \tilde{b}_{12}^{M(-h)} & \tilde{b}_{12}^{1(-h+1)} & \cdots & \tilde{b}_{12}^{Mh} \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ \tilde{b}_{1N}^{1(-h)} & \cdots & \tilde{b}_{1N}^{M(-h)} & \tilde{b}_{1N}^{1(-h+1)} & \cdots & \tilde{b}_{1N}^{Mh} \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ \tilde{b}_{N_dN}^{1(-h)} & \cdots & \tilde{b}_{N_dN}^{M(-h)} & \tilde{b}_{N_dN}^{1(-h+1)} & \cdots & \tilde{b}_{N_dN}^{Mh} \end{pmatrix} \begin{pmatrix} \Delta\tilde{r}(\mathbf{x}_1,\mathbf{h}_{-h}) \\ \vdots \\ \Delta\tilde{r}(\mathbf{x}_M,\mathbf{h}_{-h}) \\ \Delta\tilde{r}(\mathbf{x}_1,\mathbf{h}_{-h+1}) \\ \vdots \\ \Delta\tilde{r}(\mathbf{x}_M,\mathbf{h}_h) \end{pmatrix}, \quad (54)$$

where $\tilde{\mathcal{B}}(\tilde{\mathbf{r}}_0)$ is the Born operator in the extended domain. Moreover,

- $\Delta\mathbf{D} \in \mathbb{C}^{N_d \times N}$,
- $\tilde{\mathcal{B}}(\tilde{\mathbf{r}}_0) \in M_{N_d\times N, M\times N_h}(\mathbb{C})$,
- $\tilde{b}\,_{ik}^{jp} \in \mathbb{C}$, and
- $\Delta\tilde{\mathbf{r}} \in \mathbb{R}^{M\times N_h}$.

We can explicitly write the expressions for the entries of $\tilde{\mathcal{B}}(\tilde{\mathbf{r}}_0)$. $\forall i \in \{1; N_d\}, \forall k \in \{1; N\}, \forall j \in \{1; M\}, \forall p \in \{-h; h\}$, each entry $\tilde{b}\,_{ik}^{jp}$ is given by

$$\tilde{b}\,_{ik}^{jp} = F(\omega_k)\frac{2\,\omega_k^2}{m_0(\mathbf{x}_j)^{\,3}}\,G(\mathbf{x}_i,\omega_k,\mathbf{x}_j+\mathbf{h}_p;\mathbf{b},\tilde{\mathbf{r}}_0)\,G(\mathbf{x}_j-\mathbf{h}_p,\omega_k,\mathbf{x}_s;\mathbf{b},\tilde{\mathbf{r}}_0). \quad (55)$$

We have obtained the expression for the Born modeling operator in the extended subsurface offset domain (equations 53, 54, and 55), which relates a perturbation in the reflectivity model to a perturbation in the data, keeping all other parameters unchanged, and assuming a known $\mathbf{m}_0$.

Time domain expression of the Born operator

To get a better physical understanding of the Born operator, it is convenient to express our previous results in the time domain. By taking the inverse DFT of equation 43, we can express the data perturbation as the convolution in time between the Green's function and a secondary source (caused by the reflectivity perturbation). $\forall i \in \{1; N_d\}, \forall n \in \{1; N\}, \forall j \in \{1; M\}, \forall p \in \{-h; h\}$, we have

$$\Delta d(\mathbf{x}_i,n;\mathbf{b},\mathbf{r}) \;=\; \sum_{p=-h}^{h}\sum_{j=1}^{M} g(\mathbf{x}_i,n,\mathbf{x}_j+\mathbf{h}_p,0;\mathbf{b},\mathbf{r}_0) * s_{\text{sec}}(\mathbf{x}_j,\mathbf{h}_p,n). \quad (56)$$

The secondary source can be further expressed by

$$\begin{aligned} s_{\text{sec}}(\mathbf{x}_j, \mathbf{h}_p, n) &= \text{DFT}^{-1}\big[S_{\text{sec}}(\mathbf{x}_j, \mathbf{h}_p, \omega_k)\big] \\ &= \text{DFT}^{-1}\left[\frac{2F(\omega_k)\ \omega_k^2}{m_0(\mathbf{x}_j)^3}\Delta\tilde{\text{r}}(\mathbf{x}_j, \mathbf{h}_p)\ G(\mathbf{x}_j - \mathbf{h}_p, \omega_k, \mathbf{x}_s; \mathbf{b}, \mathbf{r}_0)\right] \\ &= p_{\text{src}}(\mathbf{x}_j, \mathbf{h}_p, \mathbf{x}_s, n; \mathbf{b}, \mathbf{r}_0)\ \Delta\tilde{r}(\mathbf{x}_j, \mathbf{h}_p), \end{aligned} \tag{57}$$

where,

$$p_{\text{src}}(\mathbf{x}_j, \mathbf{h}_p, \mathbf{x}_s, n; \mathbf{b}, \mathbf{r}_0) = \frac{-2}{m_0(\mathbf{x}_j)^3}\ \ddot{f}(n) * g(\mathbf{x}_j - \mathbf{h}_p, n, \mathbf{x}_s, 0; \mathbf{b}, \mathbf{r}_0). \tag{58}$$

Function $\ddot{\text{f}}$ is the discrete second order time derivative of the original source time signature. Function p_{src}, referred to as source wavefield, is a scaled version of the time convolution between the second time derivative of the source signature and the Green's function computed with the known velocity model $\mathbf{m}_0 = \mathbf{b} + \mathbf{r}_0$. $\forall i \in \{1; N_d\}, \forall n \in \{1; N\}$, the data perturbation can therefore be expressed by

$$\begin{aligned} &\Delta d(\mathbf{x}_i, n; \mathbf{b}, \mathbf{r}) = \\ &\sum_{p=-h}^{h}\sum_{j=1}^{M} g(\mathbf{x}_i, n, \mathbf{x}_j + \mathbf{h}_p, 0; \mathbf{b}, \mathbf{r}_0) * \big[p_{\text{src}}(\mathbf{x}_j, \mathbf{h}_p, \mathbf{x}_s, n; \mathbf{b}, \mathbf{r}_0)\ \Delta\tilde{r}(\mathbf{x}_j, \mathbf{h}_p)\big]. \end{aligned} \tag{59}$$

Physical interpretation of Born modeling operator

For simplicity, we focus our interpretation on the specific case of zero-subsurface offset. Let us consider the scenario where we have

- a single seismic source located at a point $\mathbf{x}_s$ at the surface,
- a subsurface location $\mathbf{x}_j$ such that $\Delta r(\mathbf{x}_j) \neq 0$, and
- a single recording location $\mathbf{x}_i$ at the surface, where we would like to compute the data perturbation $\Delta d(\mathbf{x}_i, n; \mathbf{b}, \mathbf{r})$.

The source wavefield p_{src} generated at location $\mathbf{x}_s$ is propagated into the subsurface with a known velocity model $\mathbf{m}_0$. The secondary source s_{sec} generated at $\mathbf{x}_j$, is the product of the source wavefield p_{src} with the reflectivity perturbation $\Delta r(\mathbf{x}_j)$ (equation 57). s_{sec} is nonzero if and only if the reflectivity perturbation is nonzero. In equation 59, the convolution between the secondary source and the Green's function indicates that a secondary wavefield (referred to as the scattered wavefield p_{scat}) is generated from the secondary source. Therefore, the contribution of the reflectivity perturbation $\Delta r(\mathbf{x}_j)$ to the data perturbation $\Delta d(\mathbf{x}_i, n; \mathbf{b}, \mathbf{r})$ is obtained by extracting the values of p_{scat} at location $\mathbf{x}_i$. This process is illustrated in Figure 1. Finally, in order to capture the contributions from all the reflectivity perturbations in the subsurface to an observation point $\mathbf{x}_i$, we sum over all subsurface points $\mathbf{x}_j$ (equation 59).

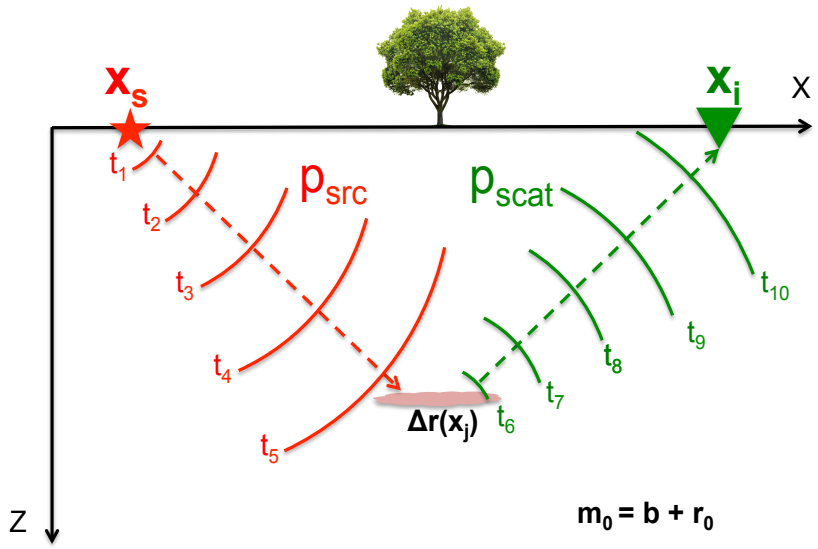


Figure 1: Schematic diagram of the Born modeling operator for one scattering point, and a known background velocity model $\mathbf{m}_0$. The source wavefield $\mathrm{p_{src}}$ (red) interacts with the reflectivity perturbation $\Delta r(\mathbf{x}_j)$ (pink), and creates a scattered wavefield $\mathrm{p_{scat}}$ (green). [**NR**] gbarnier1/. born

RTM

The RTM operator is defined as the adjoint of the Born modeling operator. From equation 54, we can obtain the adjoint $\tilde{\mathcal{R}}(\tilde{\mathbf{r}}_0)$ of operator $\tilde{\mathcal{B}}(\tilde{\mathbf{r}}_0)$ by taking its conjugate transpose matrix, $\tilde{\mathcal{R}}(\tilde{\mathbf{r}}_0) = \tilde{\mathcal{B}}(\tilde{\mathbf{r}}_0)^*$, which satisfies

$$\Delta\tilde{\mathbf{r}} = \tilde{\mathcal{R}}(\tilde{\mathbf{r}}_0)\ \Delta\mathbf{D}, \tag{60}$$

and

$$\begin{pmatrix} \Delta\tilde{r}(\mathbf{x}_1, \mathbf{h}_{-h}) \\ \vdots \\ \Delta\tilde{r}(\mathbf{x}_M, \mathbf{h}_{-h}) \\ \Delta\tilde{r}(\mathbf{x}_1, \mathbf{h}_{-h+1}) \\ \vdots \\ \Delta\tilde{r}(\mathbf{x}_M, \mathbf{h}_h) \end{pmatrix} = \begin{pmatrix} \tilde{r}_{11}^{1(-h)} & \cdots & \tilde{r}_{N_d1}^{1(-h)} & \cdots & \tilde{r}_{1N}^{1(-h)} & \cdots & \tilde{r}_{N_dN}^{1(-h)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{r}_{11}^{M(-h)} & \cdots & \tilde{r}_{N_d1}^{M(-h)} & \cdots & \tilde{r}_{1N}^{M(-h)} & \cdots & \tilde{r}_{N_dN}^{M(-h)} \\ \tilde{r}_{11}^{1(-h+1)} & \cdots & \tilde{r}_{N_d1}^{1(-h+1)} & \cdots & \tilde{r}_{1N}^{1(-h+1)} & \cdots & \tilde{r}_{N_dN}^{1(-h+1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{r}_{11}^{Mh} & \cdots & \tilde{r}_{N_d1}^{Mh} & \cdots & \tilde{r}_{1N}^{Mh} & \cdots & \tilde{r}_{N_dN}^{Mh} \end{pmatrix} \begin{pmatrix} \Delta D(\mathbf{x}_1, \omega_1) \\ \vdots \\ \Delta D(\mathbf{x}_{N_d}, \omega_1) \\ \vdots \\ \Delta D(\mathbf{x}_1, \omega_N) \\ \vdots \\ \Delta D(\mathbf{x}_{N_d}, \omega_N) \end{pmatrix}, \tag{61}$$

where

- $\Delta\tilde{\mathbf{r}} \in \mathbb{C}^{M \times N_h}$,

- $\tilde{\mathcal{R}}(\tilde{\mathbf{r}}_0) \in M_{M\times N_h, N_d\times N}(\mathbb{C})$,
- $\tilde{r}\,{}^{jp}_{ik} = (\tilde{b}\,{}^{jp}_{ik})^* \in \mathbb{C}$, and
- $\Delta\mathbf{D} \in \mathbb{R}^{M\times N_h}$.

Moreover, each entry $\tilde{r}\,{}^{jp}_{ik}$ of $\tilde{\mathcal{R}}(\tilde{\mathbf{r}}_0)$ is given by

$$\begin{aligned}\tilde{r}\,{}^{jp}_{ik} &= (\tilde{b}\,{}^{jp}_{ik})^* \\ &= F^*(\omega_k)\frac{2\,\omega_k^2}{m_0(\mathbf{x}_j)^3}\,G^*(\mathbf{x}_i,\omega_k,\mathbf{x}_j+\mathbf{h}_p;\mathbf{b},\tilde{\mathbf{r}}_0)\,G^*(\mathbf{x}_j-\mathbf{h}_p,\omega_k,\mathbf{x}_s;\mathbf{b},\tilde{\mathbf{r}}_0).\end{aligned} \tag{62}$$

Therefore, one row of equation 54 is given by

$$\Delta\tilde{r}(\mathbf{x}_j,\mathbf{h}_p) = \sum_{i=1}^{N_d}\sum_{k=1}^{N}\frac{2F^*(\omega_k)\,\omega_k^2}{m_0(\mathbf{x}_j)^3}\,G^*(\mathbf{x}_i,\omega_k,\mathbf{x}_j+\mathbf{h}_p;\mathbf{b},\tilde{\mathbf{r}}_0)\Delta D(\mathbf{x}_i,\omega_k)\,G^*(\mathbf{x}_j-\mathbf{h}_p,\omega_k,\mathbf{x}_s;\mathbf{b},\tilde{\mathbf{r}}_0). \tag{63}$$

We have obtained the expression for the RTM operator in the extended subsurface offset domain (equations 61, 62, and 63). It is the adjoint of the Born modeling operator, and it relates a perturbation in the data to a perturbation in the reflectivity model, while keeping other parameters unchanged.

Time domain expression of RTM operator

In a similar fashion as for the Born operator, we express the results obtained for the RTM operator in the time domain in order to get a better understanding of its physical meaning. $\forall j \in \{1; M\}, \forall p \in \{-h; h\}$, equation 63 can be rewritten as

$$\Delta\tilde{r}(\mathbf{x}_j,\mathbf{h}_p) = \sum_{k=1}^{N} P^*_{\text{src}}(\mathbf{x}_j,\mathbf{h}_p,\mathbf{x}_s,\omega_k;\mathbf{b},\tilde{\mathbf{r}}_0)\; P_{\text{rec}}(\mathbf{x}_j,\mathbf{h}_p,\omega_k;\mathbf{b},\tilde{\mathbf{r}}_0), \tag{64}$$

where:

- $P_{\text{src}}(\mathbf{x}_j,\mathbf{h}_p,\mathbf{x}_s,\omega_k;\mathbf{b},\tilde{\mathbf{r}}_0) = \text{DFT}\big[p_{\text{src}}(\mathbf{x}_j,\mathbf{h}_p,\mathbf{x}_s,n;\mathbf{b},\tilde{\mathbf{r}}_0)\big]$, and
- $P_{\text{rec}}(\mathbf{x}_j,\mathbf{h}_p,\omega_k;\mathbf{b},\tilde{\mathbf{r}}_0) = \sum_{i=1}^{N_d} G^*(\mathbf{x}_i,\omega_k,\mathbf{x}_j+\mathbf{h}_p;\mathbf{b},\tilde{\mathbf{r}}_0)\Delta D(\mathbf{x}_i,\omega_k)$.

P_{src} is the DFT of the source wavefield p_{src} previously defined. P_{rec} is the DFT of the *receiver wavefield*, defined as the time convolution between the data perturbation and the anti-causal Green's function g_-:

$$p_{\text{rec}}(\mathbf{x}_j, \mathbf{h}_p, n; \mathbf{b}, \tilde{\mathbf{r}}_0) = \sum_{i=1}^{N_d} g_-(\mathbf{x}_j + \mathbf{h}_p, n, \mathbf{x}_i, 0; \mathbf{b}, \tilde{\mathbf{r}}_0) * \Delta d(\mathbf{x}_i, n; \mathbf{b}, \mathbf{r}). \tag{65}$$

Therefore the receiver wavefield is the progagation *backward in time* of the data perturbation. Using equation 121 (appendix), we can show that

$$\begin{aligned} &\big[p_{\text{src}} \otimes p_{\text{rec}}\big](\mathbf{x}_j, \mathbf{h}_p, \mathbf{x}_s, 0; \mathbf{b}, \tilde{\mathbf{r}}_0) \\ &= \sum_{k=1}^{N} P^*_{\text{src}}(\mathbf{x}_j, \mathbf{h}_p, \mathbf{x}_s, \omega_k; \mathbf{b}, \tilde{\mathbf{r}}_0)\ P_{\text{rec}}(\mathbf{x}_j, \mathbf{h}_p, \omega_k; \mathbf{b}, \tilde{\mathbf{r}}_0) \\ &= \sum_{n=1}^{N} p_{\text{src}}(\mathbf{x}_j, \mathbf{h}_p, \mathbf{x}_s, n; \mathbf{b}, \tilde{\mathbf{r}}_0)\ p_{\text{rec}}(\mathbf{x}_j, \mathbf{h}_p, n; \mathbf{b}, \tilde{\mathbf{r}}_0), \end{aligned} \tag{66}$$

where the left side of equation 66 is the zero-lag time cross-correlation of the source wavefield $\mathrm{p_{src}}$ with the receiver wavefield $\mathrm{p_{rec}}$ at various subsurface locations. For any subsurface point $\mathbf{x}_j$ and for any subsurface offset $\mathbf{h}_p$, we can now express the reflectivity perturbation by

$$\Delta\tilde{r}(\mathbf{x}_j, \mathbf{h}_p) = \big[p_{\text{src}} \otimes p_{\text{rec}}\big](\mathbf{x}_j, \mathbf{h}_p, \mathbf{x}_s, 0; \mathbf{b}, \tilde{\mathbf{r}}_0), \tag{67}$$

where $\Delta\tilde{r}(\mathbf{x}_j, \mathbf{h}_p) \in \mathbb{R}$.

Physical interpretation of RTM operator

For simplicity, let us physically interpret the specific case of zero-subsurface offset, and let us consider

- a single source located at a point $\mathbf{x}_s$ at the surface, and
- a single recording location $\mathbf{x}_i$ at the surface where we have a data perturbation $\Delta D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r})$.

We apply the RTM operator in order to recover the location(s) of the reflectivity perturbation(s) that caused the data perturbation observed at the surface. The source wavefield $\mathrm{p_{src}}$ generated at location $\mathbf{x}_s$ is propagated forward in time with the known velocity model $\mathbf{m}_0$. The receiver wavefield $\mathrm{p_{rec}}$ generated by the data perturbation $\Delta D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r})$ at $\mathbf{x}_i$ is propagated backward in time with the known velocity $\mathbf{m}_0$. For each point in the subsurface, the reflectivity perturbation value is equal to the output of the cross-correlation of the two wavefields at zero-time lag. If there are locations such that both wavefields coincide at the same time, the cross-correlation output will be non-zero (assuming there is not only destructive interferences), and a reflectivity perturbation will be generated at these locations. This process is illustrated in Figure 2. Finally, to account for all the contributions coming

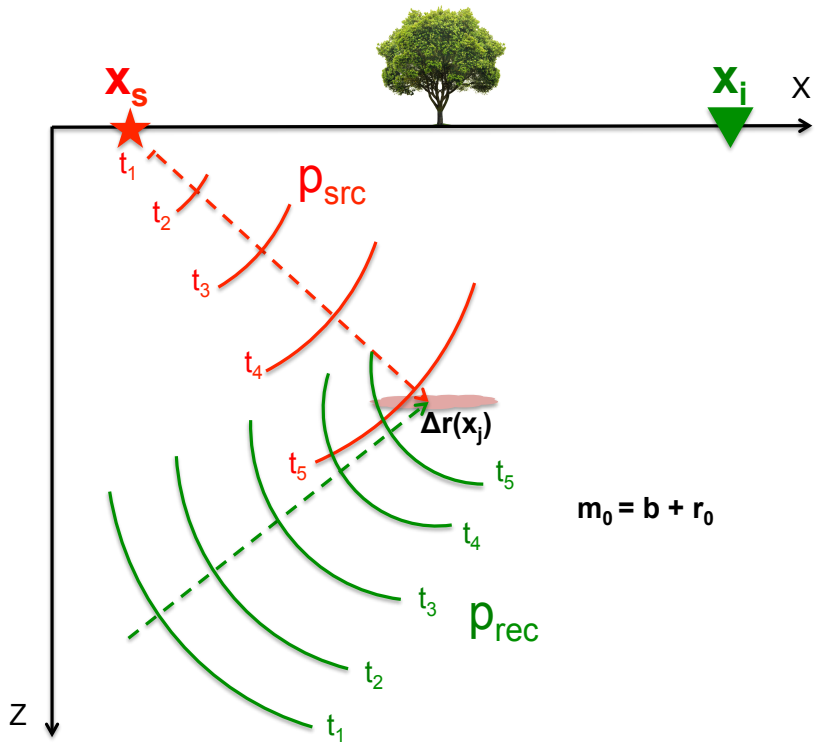


Figure 2: Schematic diagram of the RTM operator for one source located at $\mathbf{x}_s$, and one receiver located at $\mathbf{x}_i$. For all points in the subsurface, the source wavefield $\mathrm{p_{src}}$ (red) is cross-correlated at zero-time lag with the receiver wavefield $\mathrm{p_{rec}}$ (green). The receiver wavefield is propagated backward in time from the receiver location $\mathbf{x}_i$. A nonzero reflectivity perturbation will be generated at the subsurface points where the two wavefields coincide in time and space. [**NR**] gbarnier1/. rtm

from other potential observation and source locations, the reflectivity perturbations computed for each source/observation pair are summed, and a reflectivity perturbation map is generated. It is referred to as an *image*.

It is common in RTM to linearize the wave equation around a background reflectivity $\mathbf{r}_0 = \mathbf{0}$, which means that the model $\mathbf{m}_0$ used to compute the source wavefield, scattered wavefield, and receiver wavefield contains only low wavenumber components. In other words, a smooth velocity model is used to obtain the RTM image. If that is the case,

$$\Delta D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r}) \quad = \quad D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \Delta\mathbf{r}) - D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{0}), \tag{68}$$

where $D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{0})$ is the modeled data using a smooth background velocity model, which contains only direct arrivals and diving waves (no reflections). Hence, the data perturbation $\Delta D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r})$ used to compute the receiver wavefield will consist of the recorded data $D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \Delta\mathbf{r})$ from which we have removed the direct arrivals and diving waves. Though confusing, it is common to call $\Delta D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r})$ the "data."

LINEARIZATION WITH RESPECT TO THE BACKGROUND

In previous sections, we linearized the two-way wave equation with respect to the high wavenumber part of the velocity model (reflectivity), while keeping the lower wavenumber part (background) unchanged. In this section, we show how the tomographic and WEMVA operators are obtained by linearizing the Born and RTM operators with respect to the background velocity model.

Tomographic operator

The forward tomographic operator, as defined in Almomin (2013), is a linear operator that relates a perturbation in the background velocity model $\Delta\mathbf{b}$ (such that $\mathbf{b} = \mathbf{b}_0 + \Delta\mathbf{b}$) to a perturbation in the perturbation of the data $\Delta(\Delta\mathbf{D})$, while keeping the other parameters unchanged. We can symbolically write it as

$$\Delta(\Delta\mathbf{D}) = \mathbf{T}(\mathbf{b}_0, \tilde{\mathbf{r}})\Delta\mathbf{b}. \tag{69}$$

During this linearization process, we will assume that the reflectivity model $\tilde{\mathbf{r}} = \tilde{\mathbf{r}}_0 + \Delta\tilde{\mathbf{r}}$ (i.e., the image) is unchanged, and part of the operator.

Linearization of the Born operator

In order to obtain the expression for the tomographic operator, we perturb the background velocity model $\mathbf{b}$, and we express the perturbation in the perturbation of the data. We first define $\tilde{D}(\mathbf{x}_i, \omega_k, \mathbf{x}_s; \mathbf{b}, \tilde{\mathbf{r}})$ by

$$\begin{aligned}\tilde{D}(\mathbf{x}_i,\mathbf{x}_s,\omega_k;\mathbf{b},\tilde{\mathbf{r}}) &= \Delta D(\mathbf{x}_i,\mathbf{x}_s,\omega_k;\mathbf{b},\tilde{\mathbf{r}}) \\ &= D(\mathbf{x}_i,\mathbf{x}_s,\omega_k;\mathbf{b},\tilde{\mathbf{r}}) - D(\mathbf{x}_i,\mathbf{x}_s,\omega_k;\mathbf{b},\tilde{\mathbf{r}}_\mathbf{0}).\end{aligned} \tag{70}$$

We perform a multivariate first-order Taylor expansion of the data perturbation $\tilde{D}(\mathbf{x}_i,\mathbf{x}_s,\omega_k;\mathbf{b},\tilde{\mathbf{r}})$ around a known model $\mathbf{b}_0$, such that $\mathbf{b}=\mathbf{b}_0+\Delta\mathbf{b}$. $\mathbf{b}_0$ is referred to as the *background background model*, and $\Delta\mathbf{b}$ is the *background perturbation*. Moreover, throughout this process, $\tilde{\mathbf{r}}=\tilde{\mathbf{r}}_0+\Delta\tilde{\mathbf{r}}$ is assumed to be known and is kept unchanged. Assuming a small background perturbation $\Delta\mathbf{b}$, we can write

$$\begin{aligned}\Delta\tilde{D}_{ik}=\Delta\tilde{D}(\mathbf{x}_i,\mathbf{x}_s\omega_k,\mathbf{b},\tilde{\mathbf{r}}) &= \tilde{D}(\mathbf{x}_i,\mathbf{x}_s,\omega_k,;\mathbf{b},\tilde{\mathbf{r}}) - \tilde{D}(\mathbf{x}_i,\mathbf{x}_s,\omega_k;\mathbf{b}_0,\tilde{\mathbf{r}}) \\ &\approx \sum_{q=1}^{M}\frac{\partial\tilde{D}}{\partial b_q}(\mathbf{x}_i,\mathbf{x}_s,\omega_k;\mathbf{b},\tilde{\mathbf{r}})\Big|_{\mathbf{b}=\mathbf{b}_0}\Delta b(\mathbf{x}_q).\end{aligned} \tag{71}$$

Using equation 52, we have

$$\begin{aligned}\frac{\partial\tilde{D}}{\partial b_q}(\mathbf{x}_i,\mathbf{x}_s,\omega_k;\mathbf{b},\tilde{\mathbf{r}})\Big|_{\mathbf{b}=\mathbf{b}_0} &= \frac{\partial\Delta D}{\partial b_q}(\mathbf{x}_i,\mathbf{x}_s,\omega_k;\mathbf{b},\mathbf{r})\Big|_{\mathbf{b}=\mathbf{b}_0} \\ &= \sum_{j=1}^{M}\sum_{p=-h}^{h}\frac{\partial\beta_T}{\partial b_q}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{h}_p,\omega_k;\mathbf{b},\tilde{\mathbf{r}}_0)\Big|_{\mathbf{b}=\mathbf{b}_0}\alpha_T(\mathbf{x}_j,\mathbf{h}_p,\omega_k),\end{aligned} \tag{72}$$

where

- $\alpha_T(\mathbf{x}_j,\mathbf{h}_p,\omega_k)=F(\omega_k)\dfrac{2\,\omega_k^2}{m_0(\mathbf{x}_j)^3}\Delta\tilde{r}(\mathbf{x}_j,\mathbf{h}_p)$, and
- $\beta_T(\mathbf{x}_i,\mathbf{x}_j,\mathbf{h}_p,\omega_k;\mathbf{b},\tilde{\mathbf{r}}_0)=G(\mathbf{x}_i,\omega_k,\mathbf{x}_j+\mathbf{h}_p;\mathbf{b},\tilde{\mathbf{r}}_0)\,G(\mathbf{x}_j-\mathbf{h}_p,\omega_k,\mathbf{x}_s;\mathbf{b},\tilde{\mathbf{r}}_0)$.

Therefore

$$\begin{aligned}\frac{\partial\beta_T}{\partial b_q}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{h}_p,\omega_k;\mathbf{b},\tilde{\mathbf{r}}_0)\Big|_{\mathbf{b}_0} &= \frac{\partial G}{\partial b_q}(\mathbf{x}_i,\omega_k,\mathbf{x}_j+\mathbf{h}_p;\mathbf{b},\tilde{\mathbf{r}}_0)\Big|_{\mathbf{b}_0}G(\mathbf{x}_j-\mathbf{h}_p,\omega_k,\mathbf{x}_s;\mathbf{b}_0,\tilde{\mathbf{r}}_0) \\ &+ G(\mathbf{x}_i,\omega_k,\mathbf{x}_j+\mathbf{h}_p;\mathbf{b}_0,\tilde{\mathbf{r}}_0)\frac{\partial G}{\partial b_q}(\mathbf{x}_j-\mathbf{h}_p,\omega_k,\mathbf{x}_s;\mathbf{b},\tilde{\mathbf{r}}_0)\Big|_{\mathbf{b}_0}.\end{aligned}$$

We can obtain $\dfrac{\partial G}{\partial b_q}(\mathbf{x}_i,\omega_k,\mathbf{x}_j+\mathbf{h}_p;\mathbf{b},\tilde{\mathbf{r}}_0)\Big|_{\mathbf{b}_0}$ and $\dfrac{\partial G}{\partial b_q}(\mathbf{x}_j-\mathbf{h}_p,\omega_k,\mathbf{x}_s;\mathbf{b},\tilde{\mathbf{r}}_0)\Big|_{\mathbf{b}_0}$ the same way as in previous section (equation 45) by setting $\mathbf{m}_0=\mathbf{b}_0+\mathbf{r}_0$. Hence, $\forall i\in\{1;M\},\forall k\in\{1;N\},\forall j\in\{1;M\},\forall q\in\{1;M\},\forall p\in\{-h;h\}$,

$$\frac{\partial G}{\partial b_q}(\mathbf{x}_i,\omega_k,\mathbf{x}_j+\mathbf{h}_p;\mathbf{b},\tilde{\mathbf{r}}_0)\Big|_{\mathbf{b}_0} = \frac{2\,\omega_k^2}{m_0(\mathbf{x}_q)^3}\,G(\mathbf{x}_i,\omega_k,\mathbf{x}_q;\mathbf{b}_0,\tilde{\mathbf{r}}_0)\,G(\mathbf{x}_q,\omega_k,\mathbf{x}_j+\mathbf{h}_p;\mathbf{b}_0,\tilde{\mathbf{r}}_0), \tag{73}$$

and

$$\frac{\partial G}{\partial b_q}(\mathbf{x}_j-\mathbf{h}_p,\omega_k,\mathbf{x}_s;\mathbf{b},\tilde{\mathbf{r}}_0)\Big|_{\mathbf{b}_0} = \frac{2\,\omega_k^2}{m_0(\mathbf{x}_q)^3}\,G(\mathbf{x}_j-\mathbf{h}_p,\omega_k,\mathbf{x}_q;\mathbf{b}_0,\tilde{\mathbf{r}}_0)\,G(\mathbf{x}_q,\omega_k,\mathbf{x}_s;\mathbf{b}_0,\tilde{\mathbf{r}}_0). \tag{74}$$

$\forall i \in \{1;N_d\}, \forall k \in \{1;N\}$, equation 71 now becomes

$$\Delta\tilde{D}_{ik} = \Delta\tilde{D}(\mathbf{x}_i,\mathbf{x}_s,\omega_k;\mathbf{b},\tilde{\mathbf{r}}) = \sum_{q=1}^{M} T_{ik}^{\,q}\,\Delta b(\mathbf{x}_q), \tag{75}$$

with $T_{ik}^{\,q} \in \mathbb{C}$, and

$$T_{ik}^{\,q} = L_1(\mathbf{x}_i,\mathbf{x}_q,\omega_k) + L_2(\mathbf{x}_i,\mathbf{x}_q,\omega_k), \tag{76}$$

where

$$L_1(\mathbf{x}_i,\mathbf{x}_q,\omega_k) = \sum_{j=1}^{M}\sum_{p=-h}^{h} \gamma_{\,jq}^{\,k} G_0(\mathbf{x}_i,\omega_k,\mathbf{x}_q)\,G_0(\mathbf{x}_q,\omega_k,\mathbf{x}_j+\mathbf{h}_p)\Delta\tilde{r}(\mathbf{x}_j,\mathbf{h}_p)\,G_0(\mathbf{x}_j-\mathbf{h}_p,\omega_k,\mathbf{x}_s), \tag{77}$$

and

$$L_2(\mathbf{x}_i,\mathbf{x}_q,\omega_k) = \sum_{j=1}^{M}\sum_{p=-h}^{h} \gamma_{\,jq}^{\,k} G_0(\mathbf{x}_i,\omega_k,\mathbf{x}_j+\mathbf{h}_p)\Delta\tilde{r}(\mathbf{x}_j,\mathbf{h}_p)\,G_0(\mathbf{x}_j-\mathbf{h}_p,\omega_k,\mathbf{x}_q)\,G_0(\mathbf{x}_q,\omega_k,\mathbf{x}_s), \tag{78}$$

with

$$\gamma_{jq}^{k} = F(\omega_k)\frac{4\,\omega_k^4}{m_0(\mathbf{x}_j)^3\; m_0(\mathbf{x}_q)^3}. \tag{79}$$

In order to simplify notation, we used G_0 (and g_0) to denote the Green's function in the frequency domain (and time domain) computed with a velocity model $\mathbf{m}_0 = \mathbf{b}_0 + \tilde{\mathbf{r}}_0$. Equation 75 can be rewritten into the matrix form

$$\Delta\tilde{\mathbf{D}} = \begin{pmatrix} \Delta\tilde{D}_{11} \\ \Delta\tilde{D}_{21} \\ \vdots \\ \Delta\tilde{D}_{N_d1} \\ \vdots \\ \Delta\tilde{D}_{1N} \\ \vdots \\ \Delta\tilde{D}_{N_dN} \end{pmatrix} = \begin{pmatrix} T_{11}^1 & \cdots & T_{11}^M \\ T_{21}^1 & \cdots & T_{21}^M \\ \vdots & \vdots & \vdots \\ T_{N_d1}^1 & \cdots & T_{N_d1}^M \\ \vdots & \vdots & \vdots \\ T_{1N}^1 & \cdots & T_{1N}^M \\ \vdots & \vdots & \vdots \\ T_{N_dN}^1 & \cdots & T_{N_dN}^M \end{pmatrix} \begin{pmatrix} \Delta b(\mathbf{x}_1) \\ \Delta b(\mathbf{x}_2) \\ \vdots \\ \Delta b(\mathbf{x}_M) \end{pmatrix}. \tag{80}$$

Adjoint of the tomographic operator

The adjoint of the tomographic operator, as defined in Almomin (2013) is a linear operator that relates a perturbation in the perturbation of the data $\Delta(\Delta\mathbf{D})$ to a perturbation in the background velocity model $\Delta\mathbf{b}$ (such that $\mathbf{b} = \mathbf{b}_0 + \Delta\mathbf{b}$), while keeping the other parameters unchanged. It can be represented by

$$\Delta\mathbf{b} = \mathbf{T}^*(\mathbf{b}_0, \tilde{\mathbf{r}})\; \Delta(\Delta\mathbf{D}). \tag{81}$$

From equation 80, we can easily find the adjoint of the tomographic operator. We have, $\forall q \in \{1; M\}$,

$$\begin{aligned} \Delta b(\mathbf{x}_q) &= \sum_{i=1}^{N_d}\sum_{k=1}^{N}(T_{ik}^q)^*\; \Delta\tilde{D}_{ik} \\ &= \sum_{i=1}^{N_d}\sum_{k=1}^{N} L_1^*(\mathbf{x}_i, \mathbf{x}_q, \omega_k)\Delta\tilde{D}_{ik} + L_2^*(\mathbf{x}_i, \mathbf{x}_q, \omega_k)\Delta\tilde{D}_{ik}, \end{aligned} \tag{82}$$

and where $\Delta b(\mathbf{x}_q) \in \mathbb{R}$. Moreover,

$$\begin{aligned} &L_1^*(\mathbf{x}_i, \mathbf{x}_q, \omega_k)\Delta\tilde{D}_{ik} = \\ &\sum_{j=1}^{M}\sum_{p=-h}^{h}(\gamma_{jq}^{k})^*\, G_0^*(\mathbf{x}_j - \mathbf{h}_p, \omega_k, \mathbf{x}_s)\Delta\tilde{r}(\mathbf{x}_j, \mathbf{h}_p)\, G_0^*(\mathbf{x}_q, \omega_k, \mathbf{x}_j + \mathbf{h}_p)\, G_0^*(\mathbf{x}_q, \omega_k, \mathbf{x}_i)\Delta\tilde{D}_{ik}, \end{aligned} \tag{83}$$

and

$$L_2^*(\mathbf{x}_i,\mathbf{x}_q,\omega_k)\Delta\tilde{D}_{ik} = \sum_{j=1}^{M}\sum_{p=-h}^{h}(\gamma_{jq}^{k})^* G_0^*(\mathbf{x}_q,\omega_k,\mathbf{x}_s)\,G_0^*(\mathbf{x}_q,\omega_k,\mathbf{x}_j-\mathbf{h}_p)\Delta\tilde{r}(\mathbf{x}_j,\mathbf{h}_p)\,G_0^*(\mathbf{x}_j+\mathbf{h}_p,\omega_k,\mathbf{x}_i)\Delta\tilde{D}_{ik}. \tag{84}$$

We have obtained an expression for the adjoint of the tomographic operator (equations 82, 83, and 84), which relates perturbation in the perturbation of the data to perturbation in the background velocity model, while keeping the reflectivity perturbation unchanged.

Time domain expression of the adjoint of the tomographic operator

To get better insight into the adjoint of the tomographic operator, it is convenient to express our previous results in the time domain. We start by rearranging the first term of the right side of equation 82

$$\sum_{i=1}^{N_d}\sum_{k=1}^{N} L_1^*(\mathbf{x}_i,\mathbf{x}_q,\omega_k)\Delta\tilde{D}_{ik} = \sum_{k=1}^{N} P^*_{\mathrm{scat}_1}(\mathbf{x}_q,\mathbf{x}_s,\omega_k)P_{\mathrm{rec}_1}(\mathbf{x}_q,\omega_k) \tag{85}$$

where

- $P_{\mathrm{scat}_1}(\mathbf{x}_q,\mathbf{x}_s,\omega_k) = \sum_{j=1}^{M}\sum_{p=-h}^{h} P_{\mathrm{src}_1}(\mathbf{x}_j,\mathbf{h}_p,\mathbf{x}_s,\omega_k)\Delta\tilde{r}(\mathbf{x}_j,\mathbf{h}_p)\,G_0(\mathbf{x}_q,\omega_k,\mathbf{x}_j+\mathbf{h}_p)$,
- $P_{\mathrm{src}_1}(\mathbf{x}_j,\mathbf{h}_p,\mathbf{x}_s,\omega_k) = \dfrac{2\omega_k^2}{m_0(\mathbf{x}_j)^3}F(\omega_k)\,G_0(\mathbf{x}_j-\mathbf{h}_p,\omega_k,\mathbf{x}_s)$, and
- $P_{\mathrm{rec}_1}(\mathbf{x}_q,\omega_k) = \dfrac{2\omega_k^2}{m_0(\mathbf{x}_q)^3}\sum_{i=1}^{N_d} G_0^*(\mathbf{x}_q,\omega_k,\mathbf{x}_i)\Delta\tilde{D}_{ik}$.

We can express each wavefield in the time domain by taking the inverse DFT of $\mathrm{P}_{\mathrm{scat}_1}$, $\mathrm{P}_{\mathrm{src}_1}$, and $\mathrm{P}_{\mathrm{rec}_1}$. Therefore, we have

$$\begin{aligned} p_{\mathrm{scat}_1}(\mathbf{x}_q,\mathbf{x}_s,n) &= \mathrm{DFT}^{-1}\big[P_{\mathrm{scat}_1}(\mathbf{x}_q,\mathbf{x}_s,\omega_k)\big] \\ &= \sum_{j=1}^{M}\sum_{p=-h}^{h}\big[p_{\mathrm{src}_1}(\mathbf{x}_j,\mathbf{h}_p,\mathbf{x}_s,n)\Delta\tilde{r}(\mathbf{x}_j,\mathbf{h}_p)\big] * g(\mathbf{x}_q,n,\mathbf{x}_j+\mathbf{h}_p,0;\mathbf{b}_0,\tilde{\mathbf{r}}_0), \end{aligned} \tag{86}$$

$$\begin{aligned} p_{\mathrm{src}_1}(\mathbf{x}_j,\mathbf{h}_p,\mathbf{x}_s,n) &= \mathrm{DFT}^{-1}\big[P_{\mathrm{src}_1}(\mathbf{x}_j,\mathbf{h}_p,\mathbf{x}_s,\omega_k)\big] \\ &= \mathrm{DFT}^{-1}\big[\frac{2\omega_{\mathrm{k}}^2}{m_0(\mathbf{x}_j)^3}F(\omega_{\mathrm{k}})\,G_0(\mathbf{x}_j-\mathbf{h}_p,\omega_k,\mathbf{x}_s)\big] \\ &= \frac{-2}{m_0(\mathbf{x}_j)^3}\,\ddot{f}(n) * g(\mathbf{x}_j-\mathbf{h}_p,n,\mathbf{x}_s,0;\mathbf{b}_0,\tilde{\mathbf{r}}_0), \end{aligned} \tag{87}$$

$$
\begin{aligned}
p_{\mathrm{rec}_1}(\mathbf{x}_q, n) &= \mathrm{DFT}^{-1}\big[P_{\mathrm{rec}_1}(\mathbf{x}_q, \omega_k)\big] \\
&= \mathrm{DFT}^{-1}\big[\sum_{i=1}^{N_d} \frac{2\omega_{\mathrm{k}}^2}{m_0(\mathbf{x}_q)^3} G_0^*(\mathbf{x}_q, \omega_k, \mathbf{x}_i)\Delta\tilde{D}_{ik}\big] \\
&= \frac{-2}{m_0(\mathbf{x}_q)^3} \sum_{i=1}^{N_d} \ddot{g}_-(\mathbf{x}_q, n, \mathbf{x}_i, 0; \mathbf{b}_0, \tilde{\mathbf{r}}_0) * \Delta\tilde{d}(\mathbf{x}_i, n, \mathbf{x}_s; \mathbf{b}, \tilde{\mathbf{r}}).
\end{aligned} \tag{88}
$$

$\mathrm{P}_{\mathrm{rec}_1}$ is a scaled time convolution between the second time derivative of the anti-causal Green's function and the perturbation of the perturbation of the data. It is the perturbation of the perturbation of the data propagated backward in time.

Finally, using the property derived in equation 121 (appendix), we can show that

$$
\sum_{i=1}^{N_d}\sum_{k=1}^{N} L_1^*(\mathbf{x}_i, \mathbf{x}_q, \omega_k)\Delta\tilde{D}_{ik} = \big[p_{\mathrm{scat}_1} \otimes p_{\mathrm{rec}_1}\big](\mathbf{x}_q, \mathbf{x}_s, 0; \mathbf{b}_0, \tilde{\mathbf{r}}_0). \tag{89}
$$

Therefore, the first term of the right side of equation 82 is the zero-lag time cross-correlation between $\mathrm{p}_{\mathrm{scat}_1}$ and $\mathrm{p}_{\mathrm{rec}_1}$. We perform a similar analysis for the second term of the right side of equation 82

$$
\sum_{i=1}^{N_d}\sum_{k=1}^{N} L_2^*(\mathbf{x}_i, \mathbf{x}_q, \omega_k)\Delta\tilde{D}_{ik} = \big[p_{\mathrm{src}_2} \otimes p_{\mathrm{scat}_2}\big](\mathbf{x}_q, \mathbf{x}_s, 0; \mathbf{b}_0, \tilde{\mathbf{r}}_0). \tag{90}
$$

Similarly, the second term of equation 82 is the zero-lag time cross-correlation between $\mathrm{p}_{\mathrm{src}_2}$ and $\mathrm{p}_{\mathrm{scat}_2}$, which are expressed by

$$
p_{\mathrm{src}_2}(\mathbf{x}_q, \mathbf{x}_s, n) = \frac{-2}{m_0(\mathbf{x}_q)^3}\, \ddot{f}(n) * g(\mathbf{x}_q, n, \mathbf{x}_s, 0; \mathbf{b}_0, \tilde{\mathbf{r}}_0), \tag{91}
$$

$$
\begin{aligned}
&p_{\mathrm{scat}_2}(\mathbf{x}_q, \mathbf{x}_s, n) = \\
&\quad \sum_{j=1}^{M}\sum_{p=-h}^{h} \big[p_{\mathrm{rec}_2}(\mathbf{x}_j, \mathbf{h}_p, \mathbf{x}_i, n)\Delta\tilde{r}(\mathbf{x}_j, \mathbf{h}_p)\big] * g_-(\mathbf{x}_q, n, \mathbf{x}_j - \mathbf{h}_p, 0; \mathbf{b}_0, \tilde{\mathbf{r}}_0),
\end{aligned} \tag{92}
$$

and

$$
p_{\mathrm{rec}_2}(\mathbf{x}_j, \mathbf{h}_p, n) = \frac{-2}{m_0(\mathbf{x}_j)^3} \sum_{i=1}^{N_d} \ddot{g}_-(\mathbf{x}_j + \mathbf{h}_p, n, \mathbf{x}_i, 0; \mathbf{b}_0, \tilde{\mathbf{r}}_0) * \Delta\tilde{d}(\mathbf{x}_i, n, \mathbf{x}_s; \mathbf{b}, \tilde{\mathbf{r}}), \tag{93}
$$

where $\mathrm{p}_{\mathrm{rec}_2}$ corresponds to the perturbation of the perturbation of the data convolved with a scaled second time derivative of the anti-causal Green's function. Therefore, we can now explicitly rewrite equation 82 in the time domain. The background perturbation at any subsurface location $\mathbf{x}_q$, generated by a perturbation of the perturbation in the data is given by

$$\Delta b(\mathbf{x}_q) = \big[p_{\mathrm{scat}_1} \otimes p_{\mathrm{rec}_1}\big](\mathbf{x}_q, \mathbf{x}_s, 0; \mathbf{b}_0, \tilde{\mathbf{r}}_0) + \big[p_{\mathrm{src}_2} \otimes p_{\mathrm{scat}_2}\big](\mathbf{x}_q, \mathbf{x}_s, 0; \mathbf{b}_0, \tilde{\mathbf{r}}_0). \tag{94}$$

Physical interpretation of the adjoint of the tomographic operator

In order to get a physical understanding of the adjoint of the tomographic operator, we limit the study to the zero-subsurface offset case, and we consider the scenario where we have

- a single source located at a point $\mathbf{x}_s$ at the surface,
- a single recording point $\mathbf{x}_i$ (and its recorded data) at the surface,
- a known background background model $\mathbf{b}_0$,
- an unknown low wavenumber (i.e., smooth) background velocity anomaly $\Delta\mathbf{b}$, and
- a known reflectivity model $\mathbf{r}$ (i.e., an image) obtained by applying (for instance) the RTM operator using a smooth background model. That is, $\mathbf{r} = \mathbf{r}_0 + \Delta\mathbf{r}$, where $\mathbf{r}_0 = \mathbf{0}$.

Throughout this example, we assume that the reflectivity model is given and is considered a fixed part of the operator. We also assume that the background velocity anomaly is small (in magnitude) relative to the background background. We wish to find the anomaly $\Delta\mathbf{b}$ that needs to be added to the background background $\mathbf{b}_0$ to obtain the correct background model $\mathbf{b} = \mathbf{b}_0 + \Delta\mathbf{b}$. The setup of our example is illustrated in Figure 3, where a smooth background anomaly $\Delta\mathbf{b}$ is embedded into a known $\mathbf{m}_0 = \mathbf{b}_0 + \mathbf{r}$.

In order to recover the background velocity anomaly, we first need to compute the input of the adjoint of the tomographic operator. The perturbation in the perturbation of the data $\Delta\tilde{D}(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r})$ recorded at $\mathbf{x}_i$, and due to a seismic source located at $\mathbf{x}_s$, is defined in equation 71 by

$$\begin{aligned} \Delta\tilde{D}(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r}) &= \tilde{D}(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r}) - \tilde{D}(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}_0, \mathbf{r}) \\ &= D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r}) - D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{0}) - \\ &\quad \big(D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}_0, \mathbf{r}) - D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}_0, \mathbf{0})\big). \end{aligned} \tag{95}$$

As mentioned earlier, since we have chosen $\mathbf{r}_0 = \mathbf{0}$, the difference $D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r}) - D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{0})$ corresponds to the field data recorded at the observation point from which we have removed the direct arrivals and diving waves. Similarly, the difference

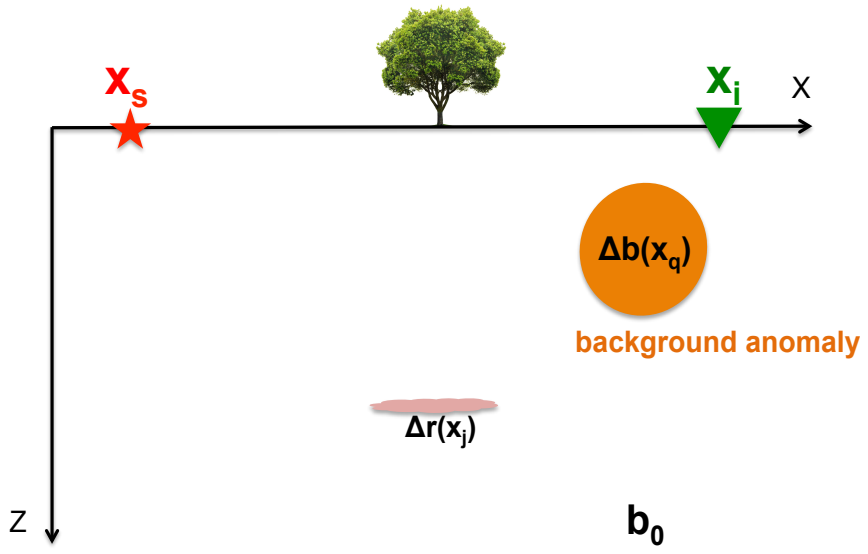


Figure 3: Schematic diagram of the true velocity model used for the experiment. A smooth and relatively small (in magnitude) background anomaly $\Delta\mathbf{b}$ (orange) is embedded into a background velocity model $\mathbf{m}_0$, such that $\mathbf{m}_0 = \mathbf{b}_0 + \mathbf{r}$. The goal is to recover the unknown anomaly $\Delta\mathbf{b}$ by applying the adjoint of the tomographic operator. [**NR**] gbarnier1/. tomosetup

$D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}_0, \mathbf{r}) - D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}_0, \mathbf{0})$ corresponds to the computed data with the background $\mathbf{b}_0$ (which does not contain the anomaly) extracted at the observation point $\mathbf{x}_i$, from which we have removed the direct arrivals and refracted waves. Assuming that we have computed $\Delta\tilde{D}(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r})$, we can now interpret the physical meaning of equation 94, starting from the first term of the right side of the equation. The sequence of schematic diagrams in Figure 4 illustrate the following process. The source wavefield $\mathrm{p}_{\mathrm{src}_1}$ is propagated forward in time from location $\mathbf{x}_s$ into the subsurface, with a known velocity model $\mathbf{m}_0 = \mathbf{b}_0 + \mathbf{r}$. In a similar fashion as for the Born modeling operator, a secondary source is created where the source wavefield interacts with a non zero reflectivity perturbation in the subsurface location $\mathbf{x}_j$. This secondary source generates a scattered wavefield $\mathrm{p}_{\mathrm{scat1}}$ (Figure 4(a)). Along with this process, the receiver wavefield $\mathrm{p}_{\mathrm{rec}_1}$ generated by the perturbation in the perturbation of the data $\Delta\tilde{D}(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r})$ at location $\mathbf{x}_i$, is propagated backward in time with the known velocity model $\mathbf{m}_0 = \mathbf{b}_0 + \mathbf{r}$ (Figure 4(b)). Finally, $\mathrm{p}_{\mathrm{scat}_1}$ and $\mathrm{p}_{\mathrm{rec}_1}$ are cross-correlated at zero-time lag at every location $\mathbf{x}_q$ in the subsurface to obtain a background velocity perturbation value $\Delta b(\mathbf{x}_q)$ (Figure 4(c)). We can clearly see that the shape of the anomaly coming from the output of the adjoint of the tomographic operator does not correspond to the one of the true anomaly. It is more elongated and less compact.

An analogous interpretation can be done for the second term of the right side of equation 94. The source wavefield $\mathrm{p}_{\mathrm{src}_2}$ is propagated forward in time from location $\mathbf{x}_s$ into the subsurface, with a known velocity model $\mathbf{m}_0 = \mathbf{b}_0 + \mathbf{r}$. The receiver wavefield $\mathrm{p}_{\mathrm{rec}_2}$ generated by the perturbation in the perturbation of the data $\Delta\tilde{D}(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r})$ at location $\mathbf{x}_i$, is propagated backward in time with the known velocity model $\mathbf{m}_0 = \mathbf{b}_0 + \mathbf{r}$. When $\mathrm{p}_{\mathrm{rec}_2}$

reaches and interacts with a non zero reflectivity perturbation in the subsurface location $\mathbf{x}_j$, a scattered wavefield $\mathrm{p}_{\mathrm{scat}_2}$ (also propagated backward in time with velocity model $\mathbf{m}_0$) is generated. Finally, $\mathrm{p}_{\mathrm{src}_2}$ and $\mathrm{p}_{\mathrm{scat}_2}$ are cross-correlated at zero-time lag at every location $\mathbf{x}_q$ in the subsurface to obtain a background velocity perturbation value $\Delta b(\mathbf{x}_q)$. The physical process associated with this second term is analogous to the one for the first term. For clarity purposes, it is not illustrated in Figure 4.

Wave-Equation Migration Velocity Analysis (WEMVA) operator

As defined in Almomin (2013), the forward WEMVA operator is an operator that relates a perturbation in the background velocity model $\Delta\mathbf{b}$ (such that $\mathbf{b} = \mathbf{b}_0 + \Delta\mathbf{b}$) to a perturbation in the perturbation of the reflectivity model $\Delta(\Delta\tilde{\mathbf{r}})$, while keeping other parameters unchanged. It is symbolically expressed by

$$\Delta(\Delta\tilde{\mathbf{r}}) = \mathbf{W}(\mathbf{b}_0, \Delta\mathbf{D})\Delta\mathbf{b}. \tag{96}$$

Throughout this process, the data perturbation $\Delta\mathbf{D}$ (expressed in equation 70) is unchanged, and is considered part of the WEMVA operator. It is analogous to the tomographic operator, but instead of performing the linearization on the Born modeling operator around a background background $\mathbf{b}_0$, it is done directly on the RTM operator itself.

Linearization of the RTM operator

In order to derive the WEMVA operator, we start with the expression of the reflectivity perturbation obtained previously for the RTM operator. The reflectivity perturbation model $\Delta\tilde{\mathbf{r}}$ obtained at all points in the subsurface was computed using a fixed background $\mathbf{b}$. Therefore, the reflectivity perturbation model $\Delta\tilde{\mathbf{r}}$ can also be considered as a function of the background. That is, $\forall j \in \{1; M\}, \forall p \in \{-h; h\}$,

$$\Delta\tilde{r}(\mathbf{x}_j, \mathbf{h}_p) = \Delta\tilde{r}(\mathbf{x}_j, \mathbf{h}_p; \mathbf{b}). \tag{97}$$

We can now perform a first-order Taylor expansion of the multivariate function $\Delta\tilde{\mathbf{r}}$ around the background background $\mathbf{b}_0$, which gives

$$\begin{aligned} \Delta(\Delta\tilde{r})(\mathbf{x}_j, \mathbf{h}_p; \mathbf{b}) &= \Delta\tilde{r}(\mathbf{x}_j, \mathbf{h}_p; \mathbf{b}) - \Delta\tilde{r}(\mathbf{x}_j, \mathbf{h}_p; \mathbf{b}_0) \\ &\approx \sum_{q=1}^{M} \frac{\partial\Delta\tilde{r}}{\partial b_q}(\mathbf{x}_j, \mathbf{h}_p; \mathbf{b})\Big|_{\mathbf{b}=\mathbf{b}_0} \Delta b(\mathbf{x}_q). \end{aligned} \tag{98}$$

We previously showed (equation 63) that for any given background model $\mathbf{b}$,

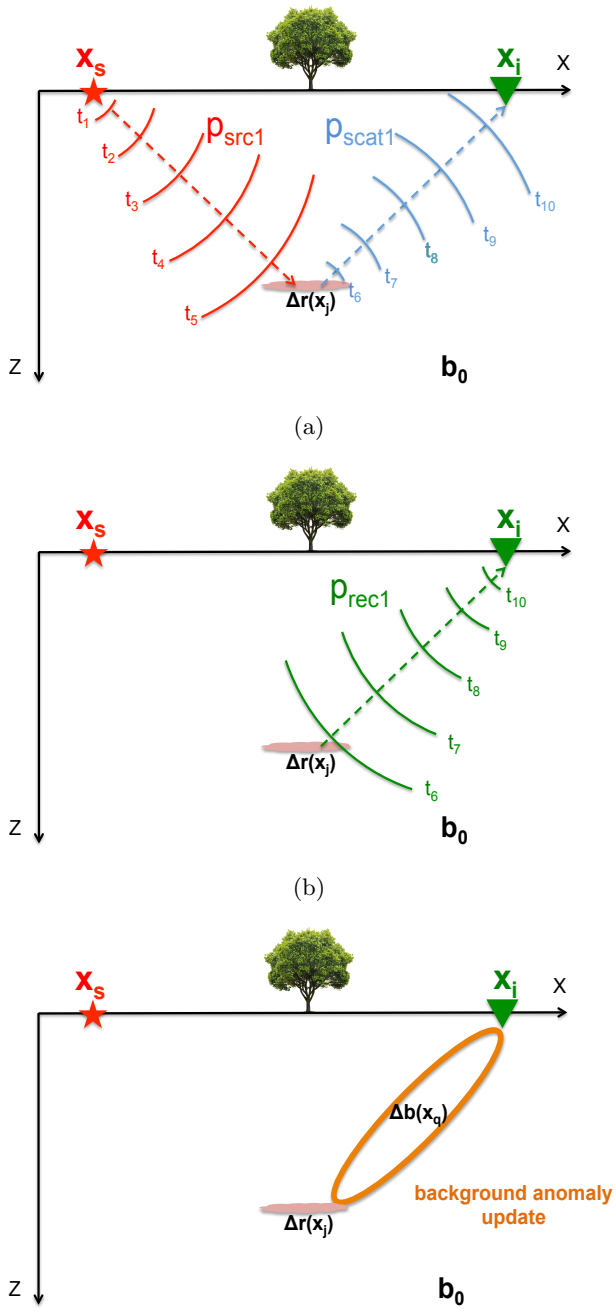


Figure 4: Sequence of schematic diagrams illustrating the adjoint of the tomographic operator applied to our example (we only show the effect of the first term in equation 94). (a) The source wavefield interacts with the reflectivity perturbation, acts as a secondary source, and generates the scattered wavefield p_{scat_1}. (b) The receiver wavefield p_{rec_1} is generated by the propagation backward in time of the perturbation of the perturbation in the data. (c) The result of the zero-time lag cross-correlation between p_{scat_1} and p_{rec_1}. [**NR**] gbarnier1/. pscat1,prec1,deltatomo

$$\Delta\tilde{r}(\mathbf{x}_j,\mathbf{h}_p;\mathbf{b}) = \sum_{i=1}^{M}\sum_{k=1}^{N}\frac{2F^*(\omega_k)\,\omega_k^2}{m_0(\mathbf{x}_j)^3}G^*(\mathbf{x}_i,\omega_k,\mathbf{x}_j+\mathbf{h}_p;\mathbf{b},\tilde{\mathbf{r}}_0)\Delta D(\mathbf{x}_i,\omega_k)\,G^*(\mathbf{x}_j-\mathbf{h}_p,\omega_k,\mathbf{x}_s;\mathbf{b},\tilde{\mathbf{r}}_0). \tag{99}$$

Now, the term $\Delta D(\mathbf{x}_i,\omega_k)$ is assumed to be known, fixed, and part of the WEMVA operator. Hence,

$$\frac{\partial\Delta\tilde{r}}{\partial b_q}(\mathbf{x}_j,\mathbf{h}_p;\mathbf{x}_s;\mathbf{b},\tilde{\mathbf{r}})\bigg|_{\mathbf{b}=\mathbf{b}_0} = \sum_{i=1}^{M}\sum_{k=1}^{N}\frac{\partial\beta_W}{\partial b_q}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{h}_p,\omega_k;\mathbf{b},\tilde{\mathbf{r}}_0)\bigg|_{\mathbf{b}=\mathbf{b}_0}\alpha_W(\mathbf{x}_j,\omega_k), \tag{100}$$

where

- $\alpha_W(\mathbf{x}_j,\omega_k) = F^*(\omega_k)\dfrac{2\,\omega_k^2}{m_0(\mathbf{x}_j)^3}\Delta D(\mathbf{x}_i,\omega_k)$, and
- $\beta_W(\mathbf{x}_i,\mathbf{x}_j,\mathbf{h}_p,\omega_k;\mathbf{b},\tilde{\mathbf{r}}_0) = G^*(\mathbf{x}_i,\omega_k,\mathbf{x}_j+\mathbf{h}_p;\mathbf{b},\tilde{\mathbf{r}}_0)\,G^*(\mathbf{x}_j-\mathbf{h}_p,\omega_k,\mathbf{x}_s;\mathbf{b},\tilde{\mathbf{r}}_0)$.

The way to evaluate $\dfrac{\partial\beta_W}{\partial b_q}(\mathbf{x}_i,\mathbf{x}_j,\mathbf{h}_p,\omega_k;\mathbf{b},\tilde{\mathbf{r}}_0)\bigg|_{\mathbf{b}=\mathbf{b}_0}$ is almost identical to the one done for the tomographic operator, and equation 99 can be rewritten in the more compact form

$$\Delta(\Delta\tilde{r})(\mathbf{x}_j,\mathbf{h}_p;\mathbf{b}) = \sum_{q=1}^{M} W_{jp}^{\;q}\,\Delta b(\mathbf{x}_q), \tag{101}$$

$\forall j\in\{1;M\},\forall p\in\{-h;h\}$. Moreover, $W_{jp}^{\;q}\in\mathbb{C}$ and

$$W_{jp}^{\;q} = L_3(\mathbf{x}_j,\mathbf{h}_p,\mathbf{x}_q) + L_4(\mathbf{x}_j,\mathbf{h}_p,\mathbf{x}_q), \tag{102}$$

where

$$L_3(\mathbf{x}_j,\mathbf{h}_p,\mathbf{x}_q) = \sum_{i=1}^{N_d}\sum_{k=1}^{N}(\gamma_{\;jq}^{\;k})^*\;G_0^*(\mathbf{x}_i,\omega_k,\mathbf{x}_q)\,G_0^*(\mathbf{x}_q,\omega_k,\mathbf{x}_j+\mathbf{h}_p)\Delta D(\mathbf{x}_i,\omega_k)\,G_0^*(\mathbf{x}_j-\mathbf{h}_p,\omega_k,\mathbf{x}_s), \tag{103}$$

and

$$L_4(\mathbf{x}_j,\mathbf{h}_p,\mathbf{x}_q) = \sum_{i=1}^{N_d}\sum_{k=1}^{N}(\gamma\,{}^{k}_{jq})^*\ G_0^*(\mathbf{x}_i,\omega_k,\mathbf{x}_j+\mathbf{h}_p)\Delta D(\mathbf{x}_i,\omega_k)\,G_0^*(\mathbf{x}_j-\mathbf{h}_p,\omega_k,\mathbf{x}_q)\,G_0^*(\mathbf{x}_q,\omega_k,\mathbf{x}_s), \tag{104}$$

with

$$\gamma\,{}^{k}_{jq} = F(\omega_k)\frac{4\,\omega_k^4}{m_0(\mathbf{x}_j)^3\ m_0(\mathbf{x}_q)^3}. \tag{105}$$

In order to simplify notations, we use G_0 and g_0 to denote the Green's functions (in the frequency and in the time domain, respectively), computed using a velocity model $\mathbf{m}_0 = \mathbf{b}_0 + \tilde{\mathbf{r}}_0$. Equation 96 can be rewritten into the matrix form

$$\Delta(\Delta\tilde{\mathbf{r}}) = \begin{pmatrix} \Delta(\Delta\tilde{r})(\mathbf{x}_1,\mathbf{h}_1;\mathbf{b}) \\ \Delta(\Delta\tilde{r})(\mathbf{x}_1,\mathbf{h}_2;\mathbf{b}) \\ \vdots \\ \Delta(\Delta\tilde{r})(\mathbf{x}_1,\mathbf{h}_{N_h};\mathbf{b}) \\ \vdots \\ \Delta(\Delta\tilde{r})(\mathbf{x}_M,\mathbf{h}_1;\mathbf{b}) \\ \vdots \\ \Delta(\Delta\tilde{r})(\mathbf{x}_M,\mathbf{h}_{N_h};\mathbf{b}) \end{pmatrix} = \begin{pmatrix} W_{11}^1 & \cdots & W_{11}^M \\ W_{12}^1 & \cdots & W_{12}^M \\ \vdots & \vdots & \vdots \\ W_{1N_h}^1 & \cdots & W_{1N_h}^M \\ \vdots & \vdots & \vdots \\ W_{M1}^1 & \cdots & W_{M1}^M \\ \vdots & \vdots & \vdots \\ W_{MN_h}^1 & \cdots & W_{MN_h}^M \end{pmatrix} \begin{pmatrix} \Delta b(\mathbf{x}_1) \\ \Delta b(\mathbf{x}_2) \\ \vdots \\ \Delta b(\mathbf{x}_M) \end{pmatrix}, \tag{106}$$

which gives us the expression for the forward WEMVA operator.

Adjoint of the WEMVA operator

The adjoint of the WEMVA operator is a linear operator that relates a perturbation in the perturbation of the reflectivity model $\Delta(\Delta\tilde{\mathbf{r}})$ to a perturbation in the background velocity model $\Delta\mathbf{b}$, while keeping other parameters unchanged (Almomin, 2013). Symbolically, it is expressed by

$$\Delta\mathbf{b} = \mathbf{W}^*(\mathbf{b}_0,\Delta\mathbf{D})\ \Delta(\Delta\tilde{\mathbf{r}}). \tag{107}$$

Equation 101 is easily adjointable, and we have $\forall q \in \{1;M\}$,

$$\begin{aligned} \Delta b(\mathbf{x}_q) &= \sum_{j=1}^{M}\sum_{p=-h}^{h}(W_{jp}^q)^*\Delta(\Delta\tilde{r})(\mathbf{x}_j,\mathbf{h}_p;\mathbf{b}) \\ &= \sum_{j=1}^{M}\sum_{p=-h}^{h}\Big(L_3^*(\mathbf{x}_j,\mathbf{h}_p,\mathbf{x}_q)+L_4^*(\mathbf{x}_j,\mathbf{h}_p,\mathbf{x}_q)\Big)\Delta(\Delta\tilde{r})(\mathbf{x}_j,\mathbf{h}_p;\mathbf{b}), \end{aligned} \tag{108}$$

where $\Delta b(\mathbf{x}_q) \in \mathbb{R}$, and

$$L_3^*(\mathbf{x}_j, \mathbf{h}_p, \mathbf{x}_q) = \sum_{i=1}^{N_d} \sum_{k=1}^{N} \gamma_{jq}^{k} \; G_0(\mathbf{x}_j - \mathbf{h}_p, \omega_k, \mathbf{x}_s) \, G_0(\mathbf{x}_q, \omega_k, \mathbf{x}_j + \mathbf{h}_p) \, G_0(\mathbf{x}_q, \omega_k, \mathbf{x}_i) \Delta D^*(\mathbf{x}_i, \omega_k), \tag{109}$$

and

$$L_4^*(\mathbf{x}_j, \mathbf{h}_p, \mathbf{x}_q) = \sum_{i=1}^{N_d} \sum_{k=1}^{N} \gamma_{jq}^{k} \; G_0(\mathbf{x}_q, \omega_k, \mathbf{x}_s) \, G_0(\mathbf{x}_q, \omega_k, \mathbf{x}_j - \mathbf{h}_p) \, G_0(\mathbf{x}_j + \mathbf{h}_p, \omega_k, \mathbf{x}_i) \Delta D^*(\mathbf{x}_i, \omega_k). \tag{110}$$

We have derived the adjoint of the WEMVA operator (equations 108, 109, and 110), which relates perturbation in the perturbation of the reflectivity $\Delta(\Delta\tilde{\mathbf{r}})$ to perturbation in the background velocity model $\Delta\mathbf{b}$, while keeping the data perturbation $\Delta\mathbf{D}$ unchanged.

Time domain expression of the adjoint of the WEMVA operator

By following a similar approach as the one done for the adjoint of the tomographic operator, we obtain the time domain expression for the background model perturbation

$$\Delta b(\mathbf{x}_q) = \big[p_{\text{scat}_3} \otimes p_{\text{rec}_3}\big](\mathbf{x}_q, \mathbf{x}_s, 0; \mathbf{b}_0, \tilde{\mathbf{r}}_0) + \big[p_{\text{src}_4} \otimes p_{\text{scat}_4}\big](\mathbf{x}_q, \mathbf{x}_s, 0; \mathbf{b}_0, \tilde{\mathbf{r}}_0), \tag{111}$$

where

$$p_{\text{scat}_3}(\mathbf{x}_q, \mathbf{x}_s, n) = \sum_{j=1}^{M} \sum_{p=-h}^{h} \big[p_{\text{src}_3}(\mathbf{x}_j, \mathbf{h}_p, \mathbf{x}_i, n) \Delta(\Delta\tilde{r})(\mathbf{x}_j, \mathbf{h}_p; \mathbf{b})\big] * g(\mathbf{x}_q, n, \mathbf{x}_j + \mathbf{h}_p, 0; \mathbf{b}_0, \tilde{\mathbf{r}}_0), \tag{112}$$

$$p_{\text{src}_3}(\mathbf{x}_j, \mathbf{h}_p, n) = \frac{-2}{m_0(\mathbf{x}_j)^3} \; \ddot{f}(n) * g(\mathbf{x}_j - \mathbf{h}_p, n, \mathbf{x}_s, 0; \mathbf{b}_0, \tilde{\mathbf{r}}_0), \tag{113}$$

$$p_{\text{rec}_3}(\mathbf{x}_q, \mathbf{x}_s, n) = \frac{-2}{m_0(\mathbf{x}_q)^3} \sum_{i=1}^{N_d} \; \ddot{g}_-(\mathbf{x}_q, n, \mathbf{x}_i, 0; \mathbf{b}_0, \tilde{\mathbf{r}}_0) * \Delta d(\mathbf{x}_i, \mathbf{x}_s, n; \mathbf{b}, \tilde{\mathbf{r}}), \tag{114}$$

and

$$p_{\text{src}_4}(\mathbf{x}_q, \mathbf{x}_s, n) \quad = \quad \frac{-2}{m_0(\mathbf{x}_q)^3}\, \ddot{f}(n) * g(\mathbf{x}_q, n, \mathbf{x}_s, 0; \mathbf{b}_0, \tilde{\mathbf{r}}_0) \tag{115}$$

$$p_{\text{scat}_4}(\mathbf{x}_q, \mathbf{x}_s, n) = \tag{116}$$
$$\sum_{j=1}^{M} \sum_{p=-h}^{h} \left[p_{\text{rec}_4}(\mathbf{x}_j, \mathbf{h}_p, \mathbf{x}_i, n) \Delta(\Delta\tilde{r})(\mathbf{x}_j, \mathbf{h}_p; \mathbf{b}) \right] * g_-(\mathbf{x}_q, n, \mathbf{x}_j - \mathbf{h}_p, 0; \mathbf{b}_0, \tilde{\mathbf{r}}_0),$$

and

$$p_{\text{rec}_4}(\mathbf{x}_j, \mathbf{h}_p, n) =$$
$$\frac{-2}{m_0(\mathbf{x}_j)^3} \sum_{i=1}^{N_d} \ddot{g}_-(\mathbf{x}_j + \mathbf{h}_p, n, \mathbf{x}_i, 0; \mathbf{b}_0, \tilde{\mathbf{r}}_0) * \Delta d(\mathbf{x}_i, \mathbf{x}_s, n; \mathbf{b}, \tilde{\mathbf{r}}). \tag{117}$$

We have derived the time domain expression of the adjoint of the WEMVA operator (equation 111). It relates a perturbation of the perturbation of the reflectivity model $\Delta(\Delta\mathbf{r})$ (which can also be interpreted as a perturbation of the image) to a perturbation of the background velocity model $\Delta\mathbf{b}$.

Physical interpretation of the adjoint of the WEMVA operator

In order to get a physical understanding of the mechanism of the adjoint of the WEMVA operator, we consider a similar scenario as for the tomographic operator where we have

- a single source located at a point $\mathbf{x}_s$ at the surface
- a single recording location $\mathbf{x}_i$ (and its recorded data) at the surface
- a known background background model $\mathbf{b}_0$
- an unknown low wavenumber velocity anomaly $\Delta\mathbf{b}$ that we would like to recover

Throughout this example, we define the data perturbation $\Delta D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r})$ as the field data recorded at observation point $\mathbf{x}_i$, from which we have removed the direct arrivals and diving waves. We will assume $\Delta D(\mathbf{x}_i, \mathbf{x}_s, \omega_k; \mathbf{b}, \mathbf{r})$ to be known and unchanged throughout this example. We wish to find the perturbation $\Delta\mathbf{b}$ that needs to be added to the background background $\mathbf{b}_0$ in order to obtain the correct background model $\mathbf{b} = \mathbf{b}_0 + \Delta\mathbf{b}$. Except for the reflectivity model, the setup of our experiment is identical to the previous example (Figure 3).
In order to recover the background velocity anomaly, we first need to compute the input of the adjoint of the WEMVA operator, which is the perturbation in the perturbation of the reflectivity, $\Delta(\Delta\tilde{\mathbf{r}})$. In this example, we assume that we have already computed a

reflectivity perturbation model $\Delta\tilde{\mathbf{r}}$, but using the incorrect background velocity model $\mathbf{b}_0$. Once $\Delta\tilde{\mathbf{r}}$ has been obtained, there are many options to compute $\Delta(\Delta\tilde{\mathbf{r}})$. We will not discuss these methods in this analysis and we assume that we have already computed $\Delta(\Delta\tilde{\mathbf{r}})$. As equations 94 and 111 indicate, the mechanism to obtain the perturbation of the background model is almost identical to the one for the adjoint of the tomographic operator. There are, however, a few variations:

- $\Delta\tilde{\mathbf{r}}$ used to be part of the tomographic operator (and kept unchanged throughout the process), and is now replaced by $\Delta(\Delta\tilde{\mathbf{r}})$,
- $\Delta(\Delta\tilde{\mathbf{r}})$ is now the "input" of the adjoint of the WEMVA operator, and
- $\Delta\tilde{\mathbf{D}}$ is replaced by $\Delta\mathbf{D}$, and is now part of the WEMVA operator, and kept unchanged during the process.

After taking into account those variations, the rest of the process to compute the update in the background velocity model is identical to the one for the adjoint of the tomographic operator, and we refer the reader to the previous section (Figure 4).

SUMMARY

We presented a detailed derivation for the Born, RTM, tomographic, and WEMVA operators by using a Born approximation and multivariate first-order Taylor expansions. We provided an expression of their forward and adjoint (both in the time and in the frequency domain), as well as a physical interpretation of their mechanism.

ACKNOWLEDGMENTS

We would like to thank the Stanford Exploration Project affiliate companies for financial support. Guillaume Barnier would like to thank Ali Almomin, Carlo Fortini, and Gustavo Alves for their unlimited patience, precious help and advice.

APPENDIX

Cross-correlation at zero time lag

We remind the reader a useful equation that relates the cross-correlation at zero-time lag of two time signals to their DFT.

- Let f and g be two discrete real time signals. We assume $\exists N \in \mathbb{N}$ such that $n \notin \{0; N-1\} \Rightarrow f(n) = g(n) = 0$.
- Let h be the time cross-correlation function between f and g. That is, $\mathrm{h} = \mathrm{f} \otimes \mathrm{g}$.
- Let F, G, and H be their respective DFT.

Therefore, $\forall n \in \{0; N-1\},\ \forall k \in \{0; N-1\}$:

$$h(n) = [f \otimes g](n) \Rightarrow H(\omega_k) = F(\omega_k)^* G(\omega_k) = F(\omega_k) G(\omega_k)^*. \tag{118}$$

The definition of the time cross-correlation h between f and g in the time domain is

$$h(n) = \sum_{k=0}^{N-1} f(k+n)\, g(k) = \sum_{k=0}^{N-1} f(k)\, g(k+n). \tag{119}$$

It is also the inverse DFT of the cross-correlation H expressed in the frequency domain,

$$\begin{aligned} h(n) = \mathrm{DFT}^{-1}[H(\omega_k)](n) &= \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} H(\omega_k)\ e^{i2\pi\ \frac{k\,n}{N}} \\ &= \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} F(\omega_k)^*\, G(\omega_k) e^{i2\pi\ \frac{k\,n}{N}} \\ &= \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} F(\omega_k)\, G(\omega_k)^* e^{i2\pi\ \frac{k\,n}{N}}. \end{aligned} \tag{120}$$

Therefore, taking the cross-correlation at zero time-lag corresponds to evaluating h at $n = 0$, which gives us

$$[f \otimes g](0) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} F(\omega_k)^*\, G(\omega_k) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} F(\omega_k)\, G(\omega_k)^* = \sum_{k=0}^{N-1} f(k)\, g(k). \tag{121}$$

REFERENCES

Almomin, A., 2013, Accurate implementation of two-way wave-equation operators: SEP-Report, **149**, 281–288.

Bob is right (per usual)

Stewart A. Levin

ABSTRACT

Bob who? —— Bob Clapp. ..... About what? —— Table lookups.
When reusable calculations are expensive, it can be advantageous to precalculate results over relevant parameter ranges and subsequently use table lookups to speed up programs that use such calculations repeatedly. Classically, sin/cos tables for fast Fourier transforms have been used to good effect. During recent development of a 3D downward continuation code, I found a range covering over an order of magnitude in the speed of sin/cos computations, with only nearest neighbor table lookup approximation outperforming Intel optimized vector routines.

INTRODUCTION

In October, 2013, I attended a ProMAX^{TM} Users Group meeting in Denver. A presenter at the meeting, Fritz Foss, now a coauthor of our upcoming SEG talk on downward continuation of Mars orbiter radar data, handed out "Martian" glasses and fascinated the audience with a description of how he was working on applying 3D poststack seismic imaging to a collection of radar tracks that had been acquired over the Martian polar icecaps. At the end of his talk, he plead for a way to avoid migrating the 90+% of zeros in his data that corresponded to the 300 km gap between the surface of Mars and the satellite's orbit. I volunteered and a collaboration ensued.

Over the months that followed, I got a slow out-of-core version working properly, but we needed it to be an order of magnitude faster to significantly outperform constant-velocity Stolt migration with the full length traces that included all the zeros. This I did achieve with resort to parallel I/O routines, but along the way I also looked at speeding up the computational downward continuation kernel. The results surprised me and may well surprise you too.

CHASING SPEEDUPS

The original downward continuation kernel was written in C, used the sin() and cos() routines from the -lm math library, and was compiled with gcc version 4.1.2 using -O3 optimization. As the version of ProMAX we were using had been mostly compiled with Intel's icc and ifort compilers, I managed to rebuild my downward continuation executable with icc version 11.1 using the -fast option. The executable did run faster, though runtime was, as anticipated, still dominated by I/O, not computational speed.

As phase shift downward continuation relies on many evaluations of $\exp(2\pi iq)$ for a limited range of shifts q, it made sense to see if table lookup would provide a performance

boost comparable to that seen in fast Fourier transform codes. (Wilcox et al. (2014)) Digging further, I found that ProMAX has a fastcos.h include file that provides macros for sin, cos and a (cos, sin) pair using table lookups. Further digging uncovered the sincos() routine in the GNU math library to produce (sin, cos) pairs.

To assess the various options, including the gcc versus the icc compiler, I wrote the test harness shown in Appendix A. The test carefully excluded parallel loop execution but biased my comparisons slightly in favor of table lookup by permitting the compiler to generate SSE (short vector) instructions to retrieve two floating point table entries, i.e. a sin/cos pair, at a go. The length of the table, that is, the number of pretabulated sin/cos pairs was set at 65536 which is 256 Kb.

Each option was timed for 2 billion random arguments resulting in the timings show in Table 1. In all timings, the minimum elapsed time of 6 trials was used, sufficient to get a reasonable estimate despite running on a computer shared with other users.

sin/cos timings	
Experiment	Δ time (seconds)
gcc -O3 table lookup	22.5
icc -fast table lookup	29.9
gcc -O3 -lm sinf() & cosf()[1]	136.9
gcc -O3 -lm sincosf()	114.8
icc -fast sinf() & cosf()	65.4
icc -fast sincosf()	65.4

Table 1: Elapsed times for $2{\times}10^9$ evaluations of single precision sin() and cos() for arguments randomly generated between $-\pi$ and π.

In reality, arguments to sin() and cos() in F-K methods are not randomly distributed but run fairly sequentially. So I retimed each option for 2 billion arguments in increasing order resulting in the timings show in Table 2.

sin/cos timings	
Experiment	Δ time (seconds)
gcc -O3 table lookup	22.1
icc -fast table lookup	22.0
gcc -O3 -lm sinf() & cosf()	118.6
gcc -O3 -lm sincosf()	96.0
icc -fast sinf() & cosf()	63.6
icc -fast sincosf()	59.4

Table 2: Elapsed times for $2{\times}10^9$ evaluations of single precision sin() and cos() for arguments linearly generated between $-\pi$ and π.

In some ways, these comparisons are unfair. The table lookups are only approximate with a maximum error equivalent to nearly 1/100th of a degree of arc. A fairer comparison

[1]These are the single precision versions.

is to use slower linear interpolation to reduce the error by another order of magnitude. Timings for this case are given in Tables 3 and 4.

sin/cos timings	
Experiment	Δ time (seconds)
gcc -O3 table lookup	35.2
icc -fast table lookup	31.4
gcc -O3 -lm sinf() & cosf()	137.1
gcc -O3 -lm sincosf()	114.3
icc -fast sinf() & cosf()	63.4
icc -fast sincosf()	63.3

Table 3: Elapsed times for 2×10^9 evaluations of single precision sin() and cos() for arguments randomly generated between $-\pi$ and π using linear table interpolation.

sin/cos timings	
Experiment	Δ time (seconds)
gcc -O3 table lookup	30.9
icc -fast table lookup	17.8
gcc -O3 -lm sinf() & cosf()	109.7
gcc -O3 -lm sincosf()	86.5
icc -fast sinf() & cosf()	57.6
icc -fast sincosf()	56.6

Table 4: Elapsed times for 2×10^9 evaluations of single precision sin() and cos() for arguments linearly generated between $-\pi$ and π using linear table interpolation.

In practice, a well written F-K code will likely arrange for arrays of arguments to sin() and cos(), not just scalars. In this setting, Intel's Math Kernel Library (MKL) provides optimized vector routines, e.g. vsSinCos(), to calculate the sin() and cos() of an array. Of course timings for this option must be compared against array-valued lookups in the sin/cos table. These results are summarized in Table 5. Timings are only shown for the sequential access pattern as, like the scalar icc timings, the times were virtually the same for random argument sequences.

sin/cos timings	
Experiment	Δ time (seconds)
icc -fast nearest neighbor table lookup	6.2
icc -fast linear interpolated table lookup	40.0
icc -fast sincosf()	16.9

Table 5: Elapsed times for 2×10^9 evaluations of single precision sin() and cos() using MKL and table lookup for arguments evaluated 1000 at a time. Random and sequential argument ordering yielded virtually identical table evaluation runtimes.

MAKING SENSE OF IT ALL

With a factor of over 20 between the fastest and slowest of the options, some explanations are in order. Certainly one observation is that the Intel icc compiler creates more optimized code that the GNU gcc compiler.

Looking at the GNU math library source code (available for download from `www.gnu.org/software/libc/download.html`), the sin() and cos() calculations involve a careful reduction of the argument to a range between $-\pi/4$ and $\pi/4$, followed by a 13^{th} or 14^{th} order polynomial evaluation involving six tabulated coefficients.

The Intel basic math library cosf() and sinf() routines call an internal library routine __libm_sse2_sincos() which, as its name implies, takes full advantage of the short vector SSE instruction and processor instruction-level parallelism. I understand that it, too, evaluates a similar polynomial.

sin() and cos() versus sincos()

The near equality in the Intel compiler and runtime results comparing sincos() with evaluating sin() and cos() separately is striking because the sincos() routine is passed memory addresses for the outputs and so would store the outputs into memory. This tells me that the Intel compiler -fast option has rolled sincos() inline into my test code, allowing the compiler to keep results in registers.

The GNU compiler and math library yield a small improvement in runtime when using the dual sincos() routine versus evaluating sin() and cos() separately. Since evaluating the sin() and cos() separately takes 118.6, each separately takes 118.6/2=59.3. Assuming the sincos() routine only evaluates one and uses sqrt() to get the other, we should expect a timing of about 59.3[sin]+ϵ[cos]+[memory stores] which says that the memory stores cost about half as much as the trigonometric evaluations.

sin() and cos() versus table lookup

The GNU compiler generates table lookups that are about the same speed as those generated by the Intel compiler. Still, this saves about a factor of 3 to 4 compared to separate or joint evaluations of sin() and cos() with their -lm math library. The penalty is a loss of accuracy with a maximum phase error of 0.0028 degrees. (For linear interpolation, the maximim phase error is 0.0000051 degrees.)

With the Intel icc compiler, table lookup was also the winner, by a factor of 2 to 3, in almost all cases. The one exception was, however, an important one: vectorized sincos() versus vectorized table lookup with linear interpolation. The reason linear interpolation is important to consider is twofold:

- The longer the trace, the more small errors in sin/cos approximation degrades the result of phase rotation.
- The 65536 entry sin/cos table I used for these tests was much larger than the 32K fast L1 cache on the machine and, indeed, was exactly the size of the slower 256K L2 cache.

Pushing for higher accuracy by using an even more refined table incurs a significant memory performance hit, whereas linear interpolation allows higher accuracy with even shorter tables.

CONCLUSIONS

So what has Bob got to do with this study? When I presented early, albeit incorrect, timing results, he added that processor speeds have continued to grow faster than memory access speeds. Where once it made excellent sense to precompute and tabulate various mathematical functions, the computational load for such a function needed to benefit from tabulation is now sufficiently high that simple trigonometric functions no longer automatically qualify. Bob Clapp emphasized that this trend is accelerating and one will soon need thousands of operations in any function to justify precomputation. In another recent seminar, Bob further made the point that the tradeoffs are even worse, by orders of magnitude, when comparing precomputed migration-inversion operator tables with on-the-fly recomputation. So, in conclusion, Bob is right (per usual).

APPENDIX A

Computer code for timing tests.

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#ifdef DOTBLFAST
#define nolinux
#endif
#include "fastcos.h"

int main(int argc, char **argv)
{
  float x=0.5f;
  float sx=0.2f, cx=-0.24f;
  float ex, fx;
  float xscale = 0.999f*M_PI;
  size_t i;
  size_t n=2000000000;

    allocfastcos();
    ex = rand()/(2000000.0f+RAND_MAX);
    for(i=0; i<n; ++i) {
       fx = 1.0f - ex;
#if defined(DORANDACCESS)
       x=(float)(xscale*(fx*(rand()/((double) RAND_MAX))+
                                       ex*(i/((double) n))));
```

```
#endif
#if defined(DOSEQACCESS)
       x=(float)(xscale*(ex*(rand()/((double) RAND_MAX))+
                                       fx*(i/((double) n))));
#endif
       ex *= (0.25f*x);
#if defined(DOTBLFAST) || defined(DOTBLSLOW)
       FASTCOSSIN(x,cx,sx);
       ex *= (cx*sx);
#endif
#if defined(DOSINCOSF)
       sincosf(x,&sx,&cx);
       ex *= (cx*sx);
#endif
#if defined(DOSINFCOSF)
       ex *= (cosf(x)*sinf(x));
#endif
    }

/* force loop evaluation by printing out final values */
    printf("x=%g cos(x)=%g sin(x)=%g\n",x,cos(x),sin(x));
#if defined(DOTBLFAST) || defined(DOTBLSLOW)
    printf("x=%g fastcos(x)=%g fastsin(x)=%g\n",x,cx,sx);
#endif
    printf("ex=%g\n",ex);

    return EXIT_SUCCESS;
}
```

REFERENCES

Wilcox, C., M. M. Strout, and J. M. Bieman, 2014, An optimization-based approach to lookup table program transformations: Journal of Software: Evolution and Process, **26**, 533–551.

Overview of the Moere Vest data set

Gustavo Alves

ABSTRACT

I present an overview of a 2D ocean bottom node survey acquired in the North Sea and made available to the Stanford Exploration Project (SEP) by Seabed Geosolutions. I describe the arrangement of the nodes, followed by the shot arrays and the data sets. The main features of this data set are four component data with offsets up to 100 kilometer, trace lengths up to 40 seconds and three different node arrangements: regular spread, long offset and microspread.

INTRODUCTION

Seismic acquisition is an ever-evolving field that is constantly being challenged by deeper and more complex targets. Illumination, geological complexities and obstructions in the survey area are only a few of the issues that have pushed the development of seismic acquisition. In marine data, the technology evolved from 2D streamer acquisition to wide azimuth and even coil-shaped sail lines. On the ocean bottom, the advent of ocean-bottom cables and ocean-bottom nodes allowed for full azimuth acquisitions, improved the repeatability in 4D surveys and introduced the recording of multi-component data in marine environments. Meunier (2011) provides a nice review on the advances in seismic acquisition.

The data provided to SEP by Seabed Geosolutions were recorded using 179 4-component ocean bottom nodes (4C OBNs) in a 2D array. The nodes were distributed in 3 different configurations, which are described here as regular spread, long offset and microspread. Three different shot-line configurations were used and the recording was sorted into 4 data sets.

Figure 1 shows the bathymetry of the survey area and the locations of the receivers. Depths vary between 1942 m (West) to 1640 m (East). The following section describes the positioning and spacing of these nodes.

NODE GEOMETRY

Regular spread

The majority of the receivers was distributed in an array of 141 nodes arranged in an East-West line, with a spacing of 250 m. The total length of this line is 35 km.

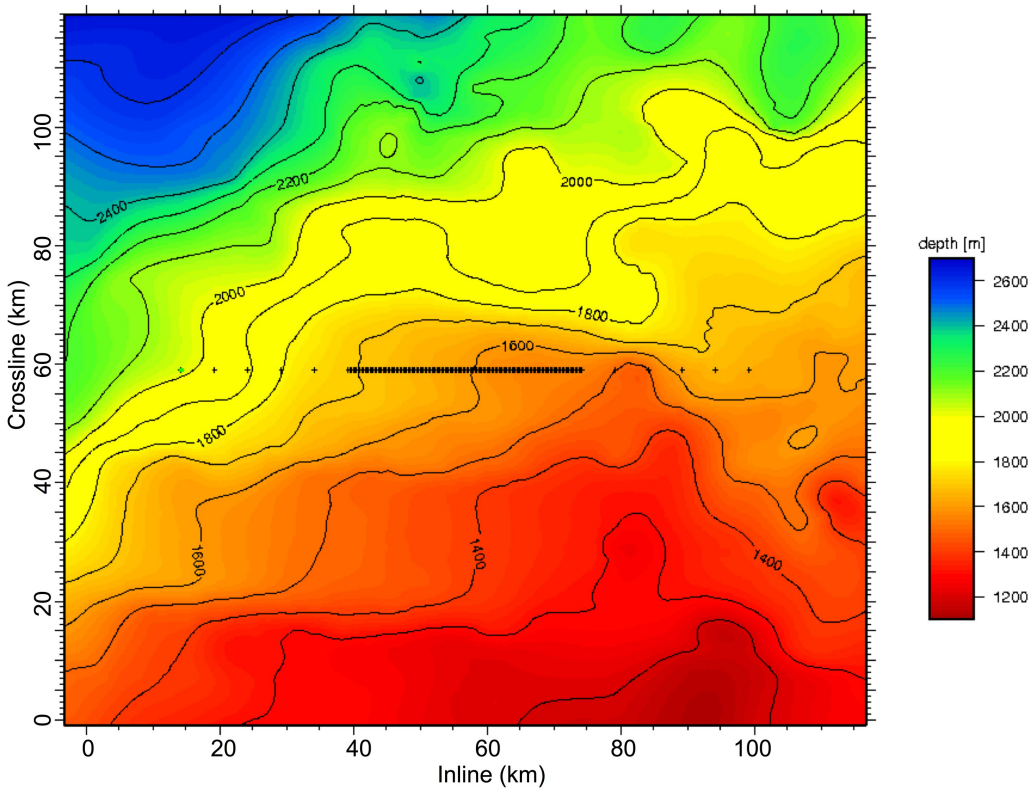


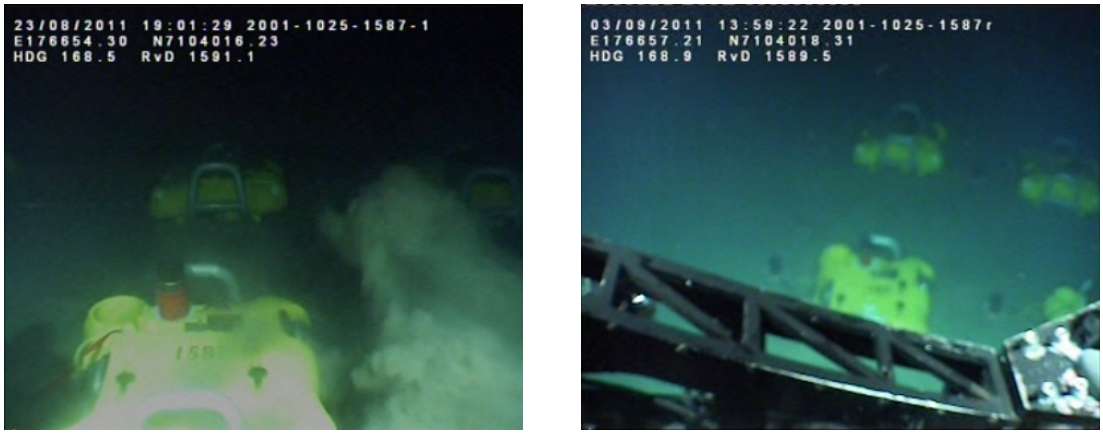
Figure 1: Bathymetry of the survey area showing the position of the receiver array. Depths vary between 1942 m (West) to 1640 m (East). Total length of the array is 85 kilometers. [**NR**] gcalves1/. bathimetry1

Long offset nodes

In order to record even longer offsets, 12 nodes were added to the 2D regular spread. They were placed at each end of the main array, 5 on each end and with a spacing of 5 km, plus an extra node at each of the two westernmost positions. These wider spaced nodes added 25 km to the length of the main receiver line on each end, increasing the total length of the array to 85 km.

Microspread

The remaining 26 nodes were placed in the middle of the array in 2 parallel lines along the main receiver line, with only a 2 meters spacing between nodes in both the inline (East-West) and crossline (North-South) directions. Figures 2(a) and 2(b) show the deployment and recovery of the nodes in this arrangement.



(a) (b)

Figure 2: Images of the (a) deployment and (b) recovery of the ocean bottom nodes in the microspread array. Note the regularity and proximity of their positions. [**NR**] gcalves1/. node1587deployment,node1587recovery

DATA SETS

Main

The data described as Main were acquired with two overlapping sail lines, each with 50 m between shots, but staggered by 25 m. The total number of shots for the Main line is 2181, with a total length of 55 km. Trace length is 15 s with trace sampling of 2 ms. A common-receiver gather of the pressure and vertical components of this data set are shown in Figures 3(a) and 3(b), respectively. The inline (radial) and crossline (transverse) components are shown in Figures 4(a) and 4(b). The gathers show, for example, the presence of refracted waves and also many orders of multiples. For the horizontal components, we can also notice the amplitude differences between the radial and transverse components.

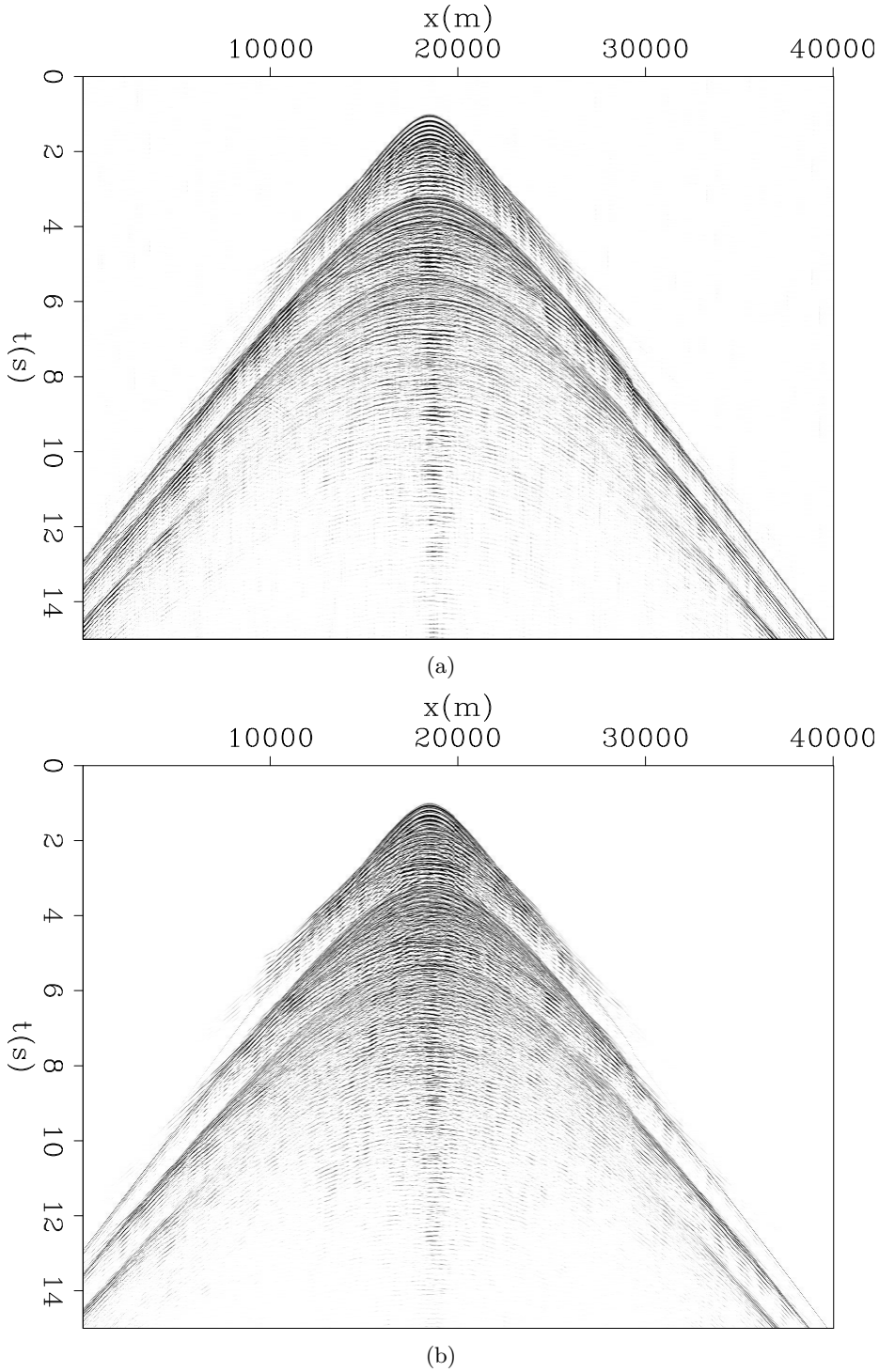


Figure 3: Common receiver gather for the (a) pressure and (b) vertical component of a node in the main data set. Distance between shots is 50 m, corresponding to a single sail line. [**ER**] gcalves1/. main-hydro,main-vert

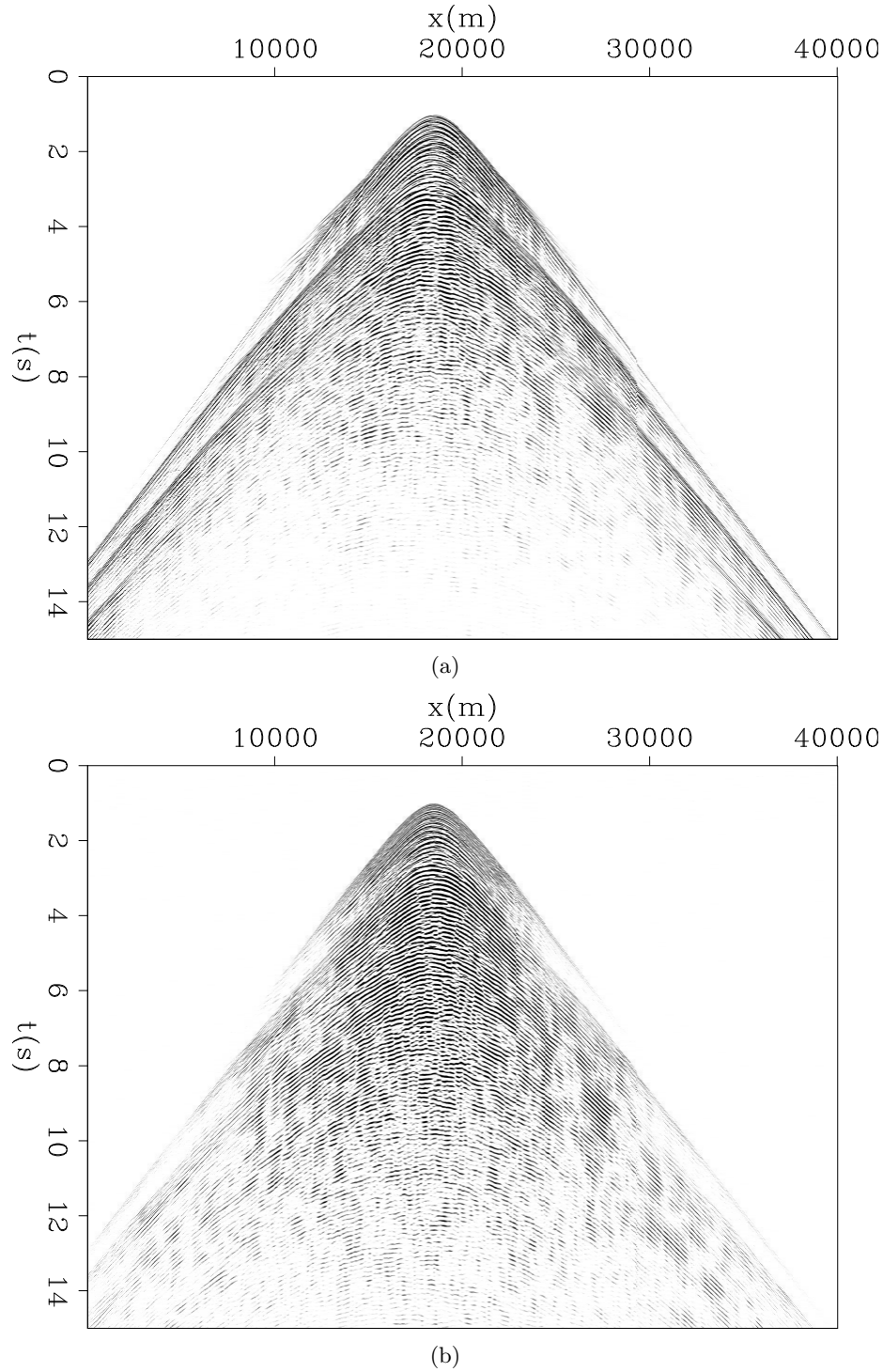


Figure 4: Common receiver gather for the (a) inline and (b) crossline horizontal components of a node in the main data set. Distance between shots is 50 m, corresponding to a single sail line. [**ER**] gcalves1/. main-y,main-x

Long

These data were recorded with the goal of acquiring long offset refraction data. There are 1140 shots, spaced by 100 m. Traces are 40 s long with 2 ms sampling. The maximum offset is 75 km for the regular spread and up to 100 km if the long offset nodes are included. Figures 5 and 6 show the hydrophone and vertical components for a node in the long data set.

Swath

This data set was acquired with 6 parallel sail lines in the East-West direction. Each sail line had 2 sources in a flip-flop arrangement, resulting in a 50 m by 50 m shot carpet, that is 600 m wide and 55 km long.

Continuous

Seabed also made available the continuous record of each node during the survey period. This period corresponds to roughly 10 days of continuous data, providing an interesting resource for passive seismic studies. Due to the different times for placement and removal of each receiver, the start and end times for the recordings vary, with an overlap of about 7 days when all receivers were in position and recording.

ACKNOWLEDGMENT

The author would like to thank Seabed Geosolutions for releasing the field data and for supporting the data handover by Paul Milligan and Shuki Ronen. The author would also like to thank Petrobras for support of his PhD.

REFERENCES

Meunier, J., 2011, Seismic acquisition from yesterday to tomorrow: Society of Exploration Geophysicists.

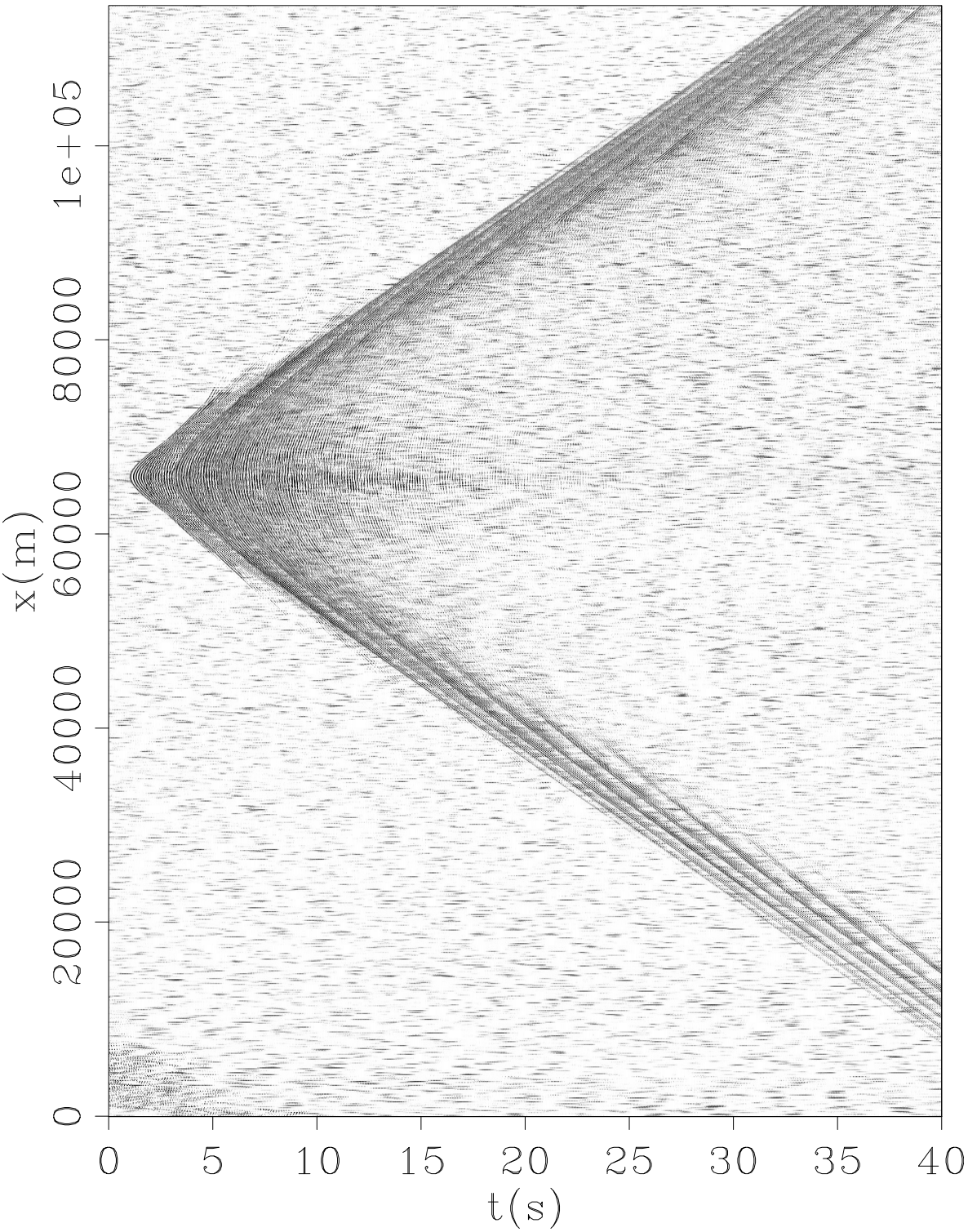


Figure 5: Common receiver gather for the pressure component of a node in the long data set. Distance between shots is 100 m. [**ER**] gcalves1/. long-hydro

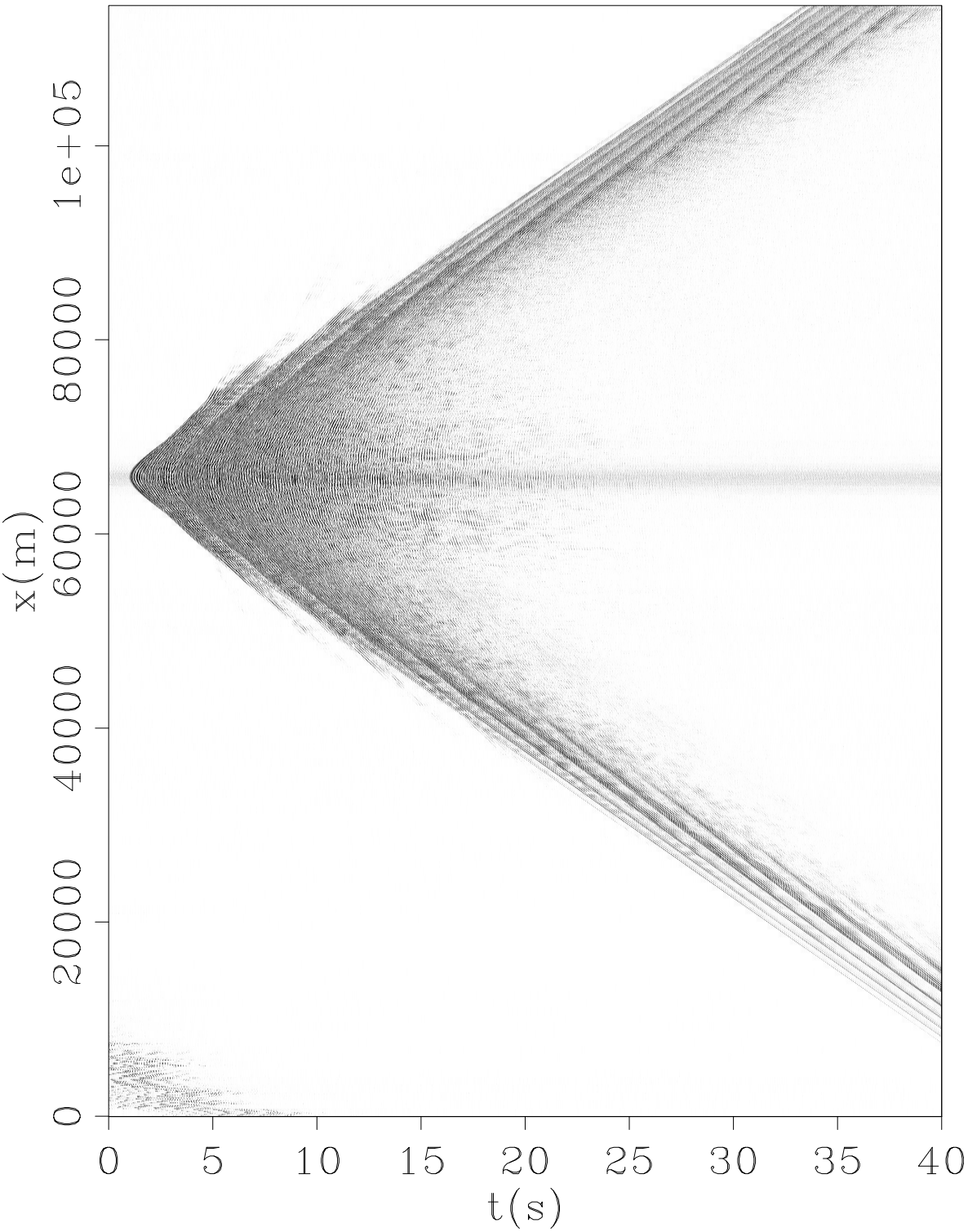


Figure 6: Common receiver gather for the vertical component of a node in the long data set. Distance between shots is 100 m. [**ER**] gcalves1/. long-vert

Analysis of Moere Vest OBNs as continuous data

Sjoerd de Ridder

ABSTRACT

The Moere Vest Ocean Bottom Node Survey was acquired for the purpose of exploration with controlled seismic source. However, the nodes recorded continuously. This paper will explore the use of the ambient seismic field for the purpose of creating low frequency virtual sources by seismic interferometry. I find that this OBN dataset contains abundant energy at microseism frequencies that yield virtual seismic sources by cross-correlations. The cross-correlation gathers contain dispersive interface waves and events with a hyperbolic move out.

INTRODUCTION

Ocean Bottom Nodes (OBNs) are placed on the seafloor and record continuously for the duration of the deployment. A vessel carrying a seismic source will sail over the nodes and shoot a seismic survey. The data is downloaded after the nodes are retrieved. The continuous recordings are typically cut into shorter records starting at the timestamp of each shot.

However, here we study the data as continuous recordings because the nodes recordings contain a wealth of seismic energy besides the controlled source shooting. Energy in the frequency range from 0.1 to 2.0 Hz is referred to as microseisms (Longuet-Higgins, 1950). The aim of this paper is to show that these data contains sufficient microseism energy to be of interest for seismic interferometry.

Seismic interferometry is a technique that aims to create virtual seismic sources by cross-correlations of ambient seismic energy (Aki, 1957; Claerbout, 1968; Lobkis and Weaver, 2001; Weaver and Lobkis, 2002; Wapenaar, 2003, 2004; Wapenaar and Fokkema, 2006). Under certain requirements on the characteristics of the ambient seismic field, a cross-correlation of recordings made at two stations results in a signal that is proportional to the Green's function of wave propagation between those two stations.

First, I briefly survey all the energy recorded by computing a spectragram that shows how the spectrum of the recordings vary as a function of time. Second, I perform cross-correlations of the microseism energy recorded in the hydrophones of all 179 stations.

MOERE VEST OBN RECORDINGS AS CONTINUOUS DATA

This dataset is comprised of 179 ocean bottom stations in a linear array spanning almost 70 km, see Alves (2014). However, the seismic shooting extended beyond the array to record long offsets. Figure 1 contains a map with the location of the Moere Vest OBN survey in the Norwegian Sea and a bathymetry map of the Moere Vest area with the station locations.

The station spacing of the center part of the line with 141 stations is 250 m. An additional 5 stations on either side of the line with spacing 5000 m provide for recordings at long offsets, each end had an extra station. At the center of the line there were 26 stations in a microspread with spacing as short as 2 m.

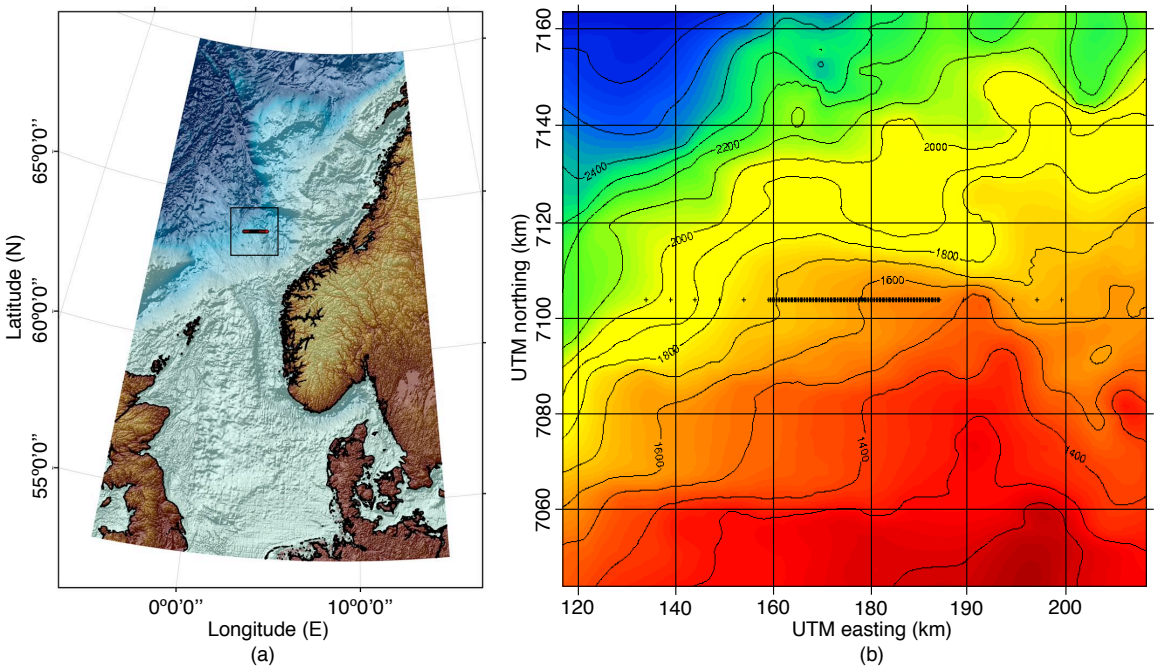


Figure 1: a) Map showing the location of the Moere Vest OBN survey. b) Bathymetry map of the Moere Vest area showing the stations in black crosses. Contour lines are drawn at 100 m interval. [**NR**] sjoerd2/. maps

The stations are not deployed simultaneously. The stations are turned on aboard the vessel and an unmanned Remotely Operated underwater Vehicle (ROV) flies a few stations at a time to the seabed for deployment. Figure 2 shows the deployment times for each node. The node numbering is from west to east. Although the recording from one node ended prematurely, all stations were recording simultaneously for a duration of about 7 days.

I investigate the frequency content of the recordings by computing a spectragram for three stations; the first station to be deployed; last station to be deployed; and a station from the microspread. The frequency spectrum of the hydrophone recordings is computed in non-overlapping 2-minute long windows, and each spectrum is displayed side by side to show how the spectrum varies as a function of time (Figures 3a-c).

The low frequencies of the Moere Vest OBN recording (between 0.1 and 2.0 Hz) contain energy that is continuously present. This energy is referred to as microseism noise and is weather generated. Generally, we expect two energy bands in this regime (the single- and double-frequency microseism peaks). We observe these two energy bands during August 23^{rd} to 24^{th}. However for the majority of the recordings we observe only one broad energy band. It remains to be investigated whether we can distinguish a single and double microseism peak, or if the energy at the upper and lower ends of the microseism frequency range is composed of waves from different propagation regimes or wave types.

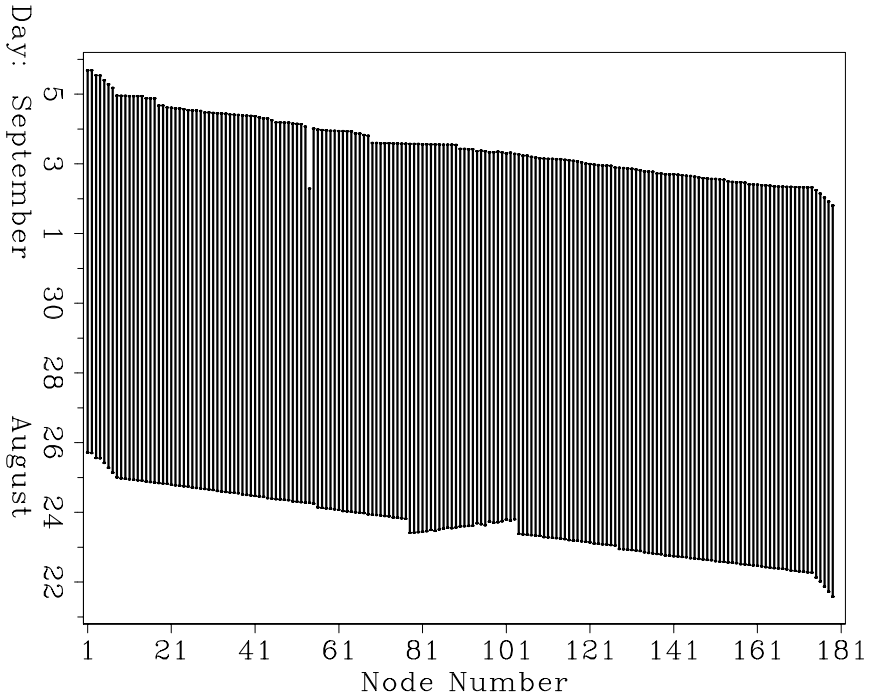


Figure 2: Deployment times of each node. All stations were recording simultaneously for a duration of about 7 days. [**ER**] sjoerd2/. deploytimes

At the higher frequencies of the Moere Vest OBN recording (above 2 Hz), we observe several controlled source seismic surveys. These are most easily identified by observing that shooting happens in episodes. During August 26^{th} to 28^{th} a few long shot lines were recorded for the whole line. Later, between August 31^{st} to September 2^{nd} six shorter shot lines were recorded along the microspread.

In order to cross-correlate the recordings made at different nodes, they need to be synchronized in time. The best approach is to add zeros before and after each recording to ensure that each trace is equally long. The first 2 hours and last 2 hours of each recording are discarded, because the nodes were not yet positioned on the sea floor. With a Hann window we taper an additional 2 hours at the beginning and end of the recordings. Figure 4 contains a gather of all hydrophone recordings bandpass filtered between 0.5 and 1.5 Hz, with zeros added before and after each trace, to align them in time.

The nodes in the center of the gather in Figure 4 all appear similar. These nodes are part of the microspread. Those nodes are so close together that the waveforms should appear very similar between nearby nodes. While at early times the arrivals in the microspread appear aligned, at a later times certain recordings in the microspread become misaligned with their neighboring recordings. This indicates that the internal clocks of the nodes do not stay synchronized, a phenomena referred to as clock drift. Hatchell and Mehta (2010) used seismic interferometry to estimate clock drifts and positioning errors.

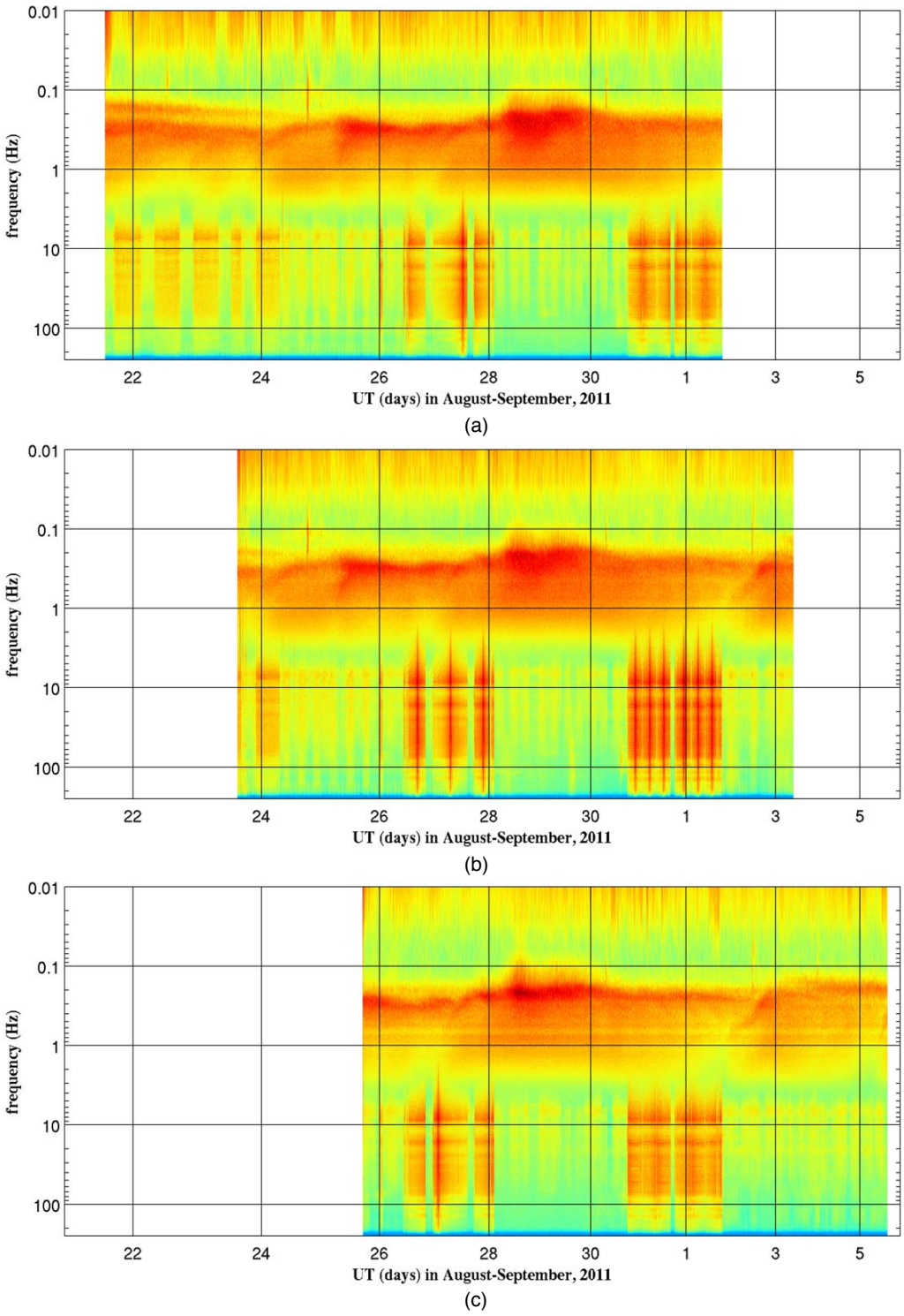


Figure 3: Spectragrams showing how the frequency content of the hydrophone recordings vary as a function of time. a) The first station to be deployed, b) a station from the microspread, c) the last station to be deployed. [**ER**] sjoerd2/. spgram

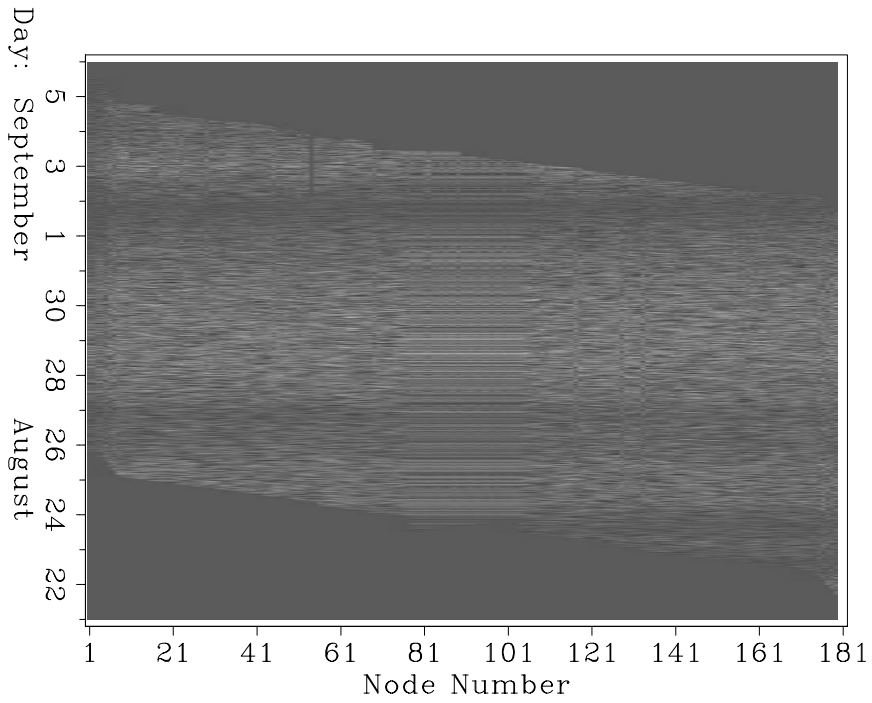


Figure 4: A gather of all the recordings synchronized with their deployment times. The gather is bandpass filtered for microseism energy between 0.5 and 1.5 Hz. The microspread in the middle shows clock drift: the nodes were better synchronized in August and less in September. [**ER**] sjoerd2/. moerevestHbp-1p5hz

CROSS-CORRELATIONS OF MICROSEISM NOISE

Seismic interferometry is a theory that describes how interstation Green's functions can be retrieved from ambient seismic recordings by a cross-correlation. The hydrophone recordings from all stations were aligned in time and padded with zeros (see previous section). For the limited purposed of this paper I ignore the clock drift. However, a more in-depth study using seismic interferometry on these data has to include a correction of the data for the clock drift prior to cross-correlations.

I elect the trace from one station as a master trace and cross-correlate the aligned and zero padded data for all 179 traces with this master trace. This effectively retrieves a gather as if there was a source at the master station. I repeat this procedure for 12 stations along the line. All cross-correlation gathers are shown in Figure 5.

The correlation gathers generally show more energy at negative correlation lags for stations on the east side of the master station and at positive correlation lags for stations on the west side of the master station. This indicates that the dominant direction of wave propagation in the ambient field was towards the west (away from the shallower waters and Norwegian coast line). Figure 6 contains a regularized offset gather of all the gathers from Figure 5.

Studying Figures 5 and 6, it is apparent that they contain dispersive interface waves traveling with a group velocity of about 1333 m/s. But there is also energy arriving with an apparent hyperbolic move out as a function of offset. These events intersect with the time-axis in the first 10 seconds. These hyporbolas are already visible in the cross-correlation gathers of individual virtual seismic sources in Figure 5.

CONCLUSIONS

This paper revealed that during the Moere Vest OBN survey there is abundant energy in the microseism frequency range 0.1 to 2.0 Hz. From cross correlations we find that the microseism energy is propagating predominantly westward. In the cross-correlation gathers we find dispersive interface waves with a group velocity of about 1333 m/s. But more interestingly we find events with an apparent hyperbolic move out. A more extensive analysis should reveal what kind of waves they are and where they originate from.

ACKNOWLEDGMENTS

I thank Seabed Geosolutions for releasing the data to Stanford and assistance in handing it over. I thank Biondo Biondi and Jon Claerbout from the Stanford Exploration Project; Shuki Ronen, Paul Milligan and Larry Wagner from Seabed Geosolutions for helpful discussions and suggestions; Gustavo Alves and Ohad Barak for close collaborations on the dataset and doing the hard work of handling the data formats. I thank Chris Castillo for generating the map in Figure 1a. This work was carried out during my visit at Stanford University during July and August of 2014. I thank Chevron's Center of Research Excellence (CORE) at Stanford for supporting my visit to Stanford.

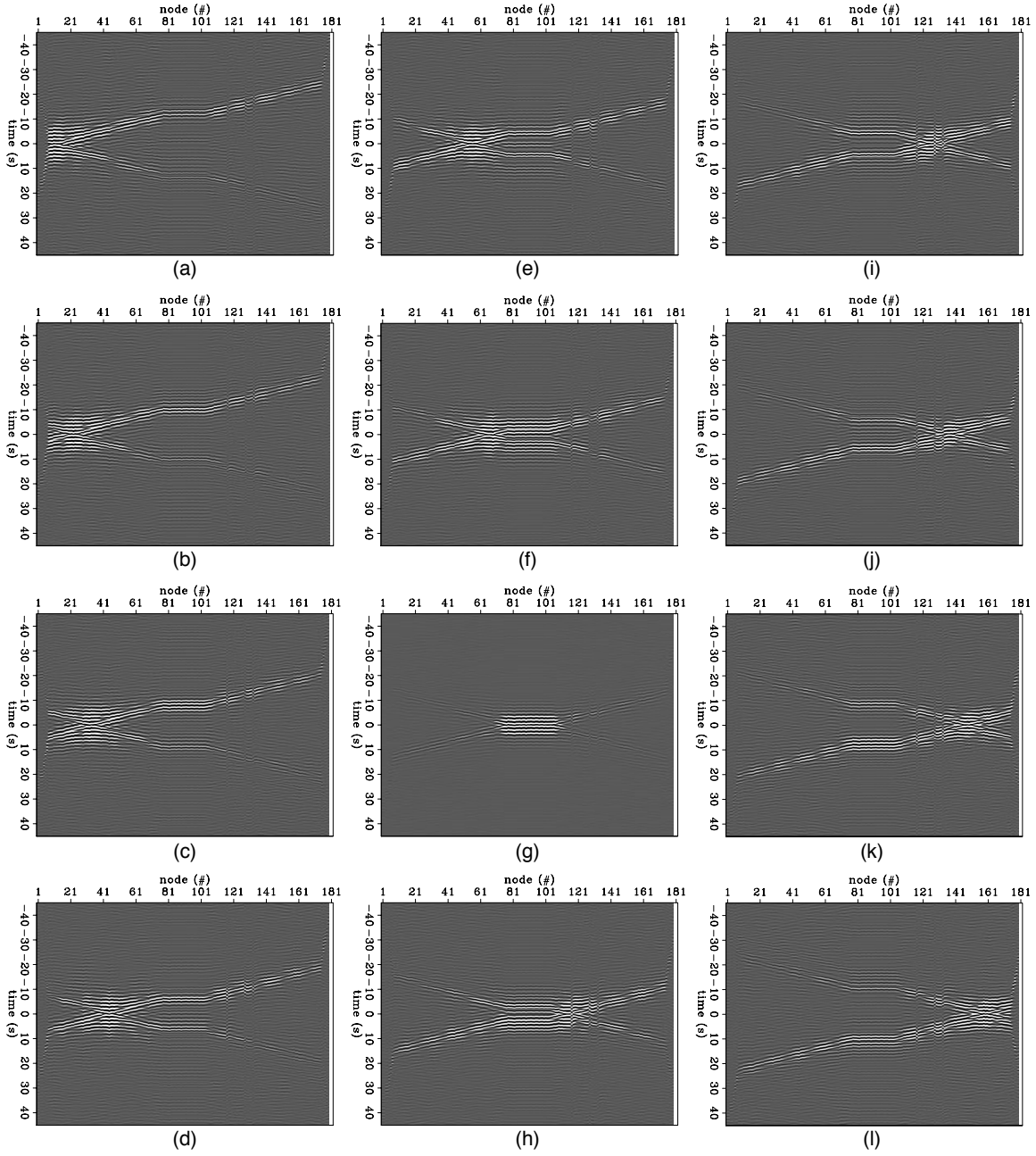


Figure 5: Cross-correlation gathers,sorted from west to east, of the data recorded between different nodes. a-l) Each contains a gather for which a different master station is turned into a virtual seismic source. [**ER**] sjoerd2/. si

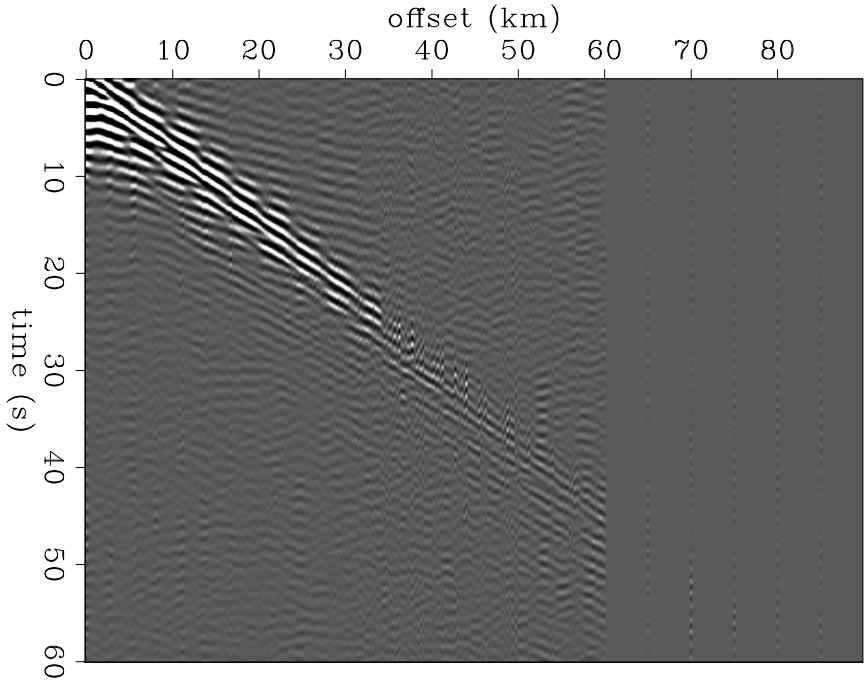


Figure 6: Offset sorted gather of all cross-correlations of Figure 5. [**ER**] sjoerd2/. sis3

REFERENCES

Aki, K., 1957, Space and time spectra of stationary stochastic waves, with special reference to microtremors: Bulletin of the Earthquake Research Institute, **35**, 415–456.

Alves, G., 2014, Overview of the moere vest data set.

Claerbout, J. F., 1968, Synthesis of a layered medium from its acoustic transmission response: Geophysics, **33**, 264–269.

Hatchell, P. J., and K. J. Mehta, 2010, Passive interferometry of deepwater ocean bottom seismic data: 72^{st} Conference and Exhibition, EAGE, Extended Abstracts, G012.

Lobkis, O. I., and R. L. Weaver, 2001, On the emergence of the Green's function in the correlations of a diffuse field: Journal of the Acoustical Society of America, **110**, 3011–3017.

Longuet-Higgins, M. S., 1950, A theory of the origin of microseisms: Philosophical Transactions of the Royal Society of London A, **243**, 1–35.

Wapenaar, K., 2003, Synthesis of an inhomogeneous medium from its acoustic transmission response: Geophysics, **68**, 1756–1759.

———, 2004, Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation: Physics Review Letters, **93**, 254301–1–254301–4.

Wapenaar, K., and J. Fokkema, 2006, Green's function representations for seismic interferometry: Geophysics, **71**, no. 4, SI33–SI46.

Weaver, R. L., and O. I. Lobkis, 2002, On the emergence of the Green's function in the correlations of a diffuse field: pulse-echo using thermal phonons: Ultrasonics, **40**, 435–439.

APPENDIX: MICROSEISM ENERGY VERSUS CONTROLLED-SOURCE SEISMIC ENERGY

In this appendix I show the dominance of the microseism energy in the low frequencies of the recordings. I perform a series of bandpass filters on a common receiver gather of the hydrophone recordings and the vertical component geophone recordings (Figures B-1 and B-2). The bandpass is implemented in the frequency domain with Hann window with a flat center.

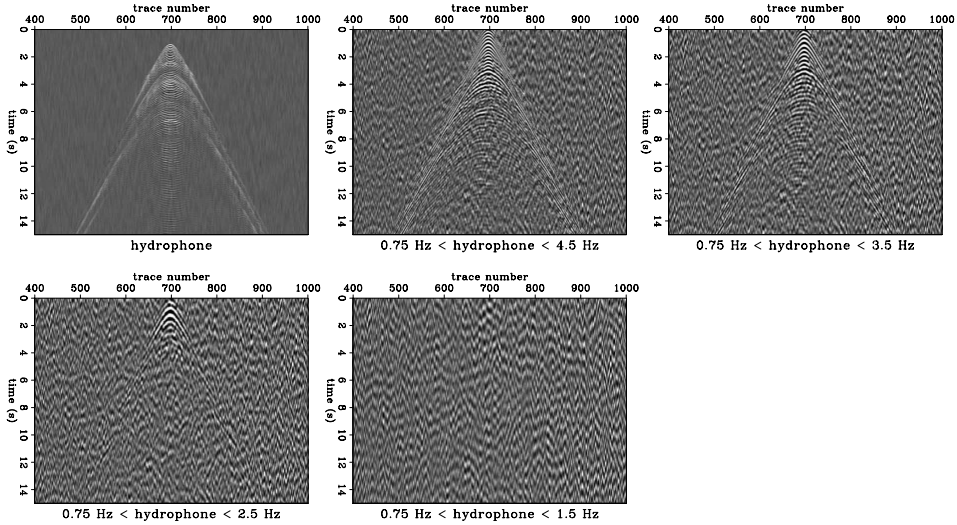


Figure B-1: Hydrophone common receiver gather: the original (upper left) and a series of bandpass filtered gathers showing the low frequency content. [**ER**] sjoerd2/. bandpassesH

Below 5 Hz there is little source energy left. However, the remaining low-frequency seismic source energy may contain reflections and refractions of the dominant and high-contrast geological formations in the subsurface. From the spectragrams in Figure 3 we see that the energy of the controlled source shooting and the energy of the microseism noise overlap in the frequency range 1 to 2 Hz. With two bandpass filters (flat center with Hann window sides) tied at respectively 0.75-1.0-1.25-1.75 Hz and 1.75-2.0-2.25-2.75 Hz, we can see that 1.75 is a good upper limit for a bandpass filter to separate the frequency ranges in which the microseism noise or the controlled source seismic energy is dominant (Figure B-3).

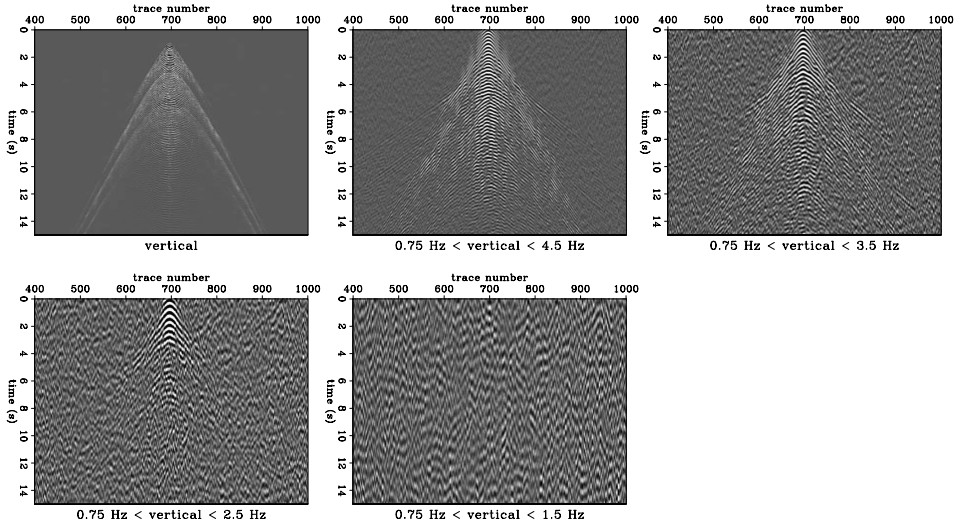


Figure B-2: Vertical geophone common receiver gather: the original (upper left) and a series of bandpass filtered gathers showing the low frequency content. [**ER**] sjoerd2/. bandpassesV

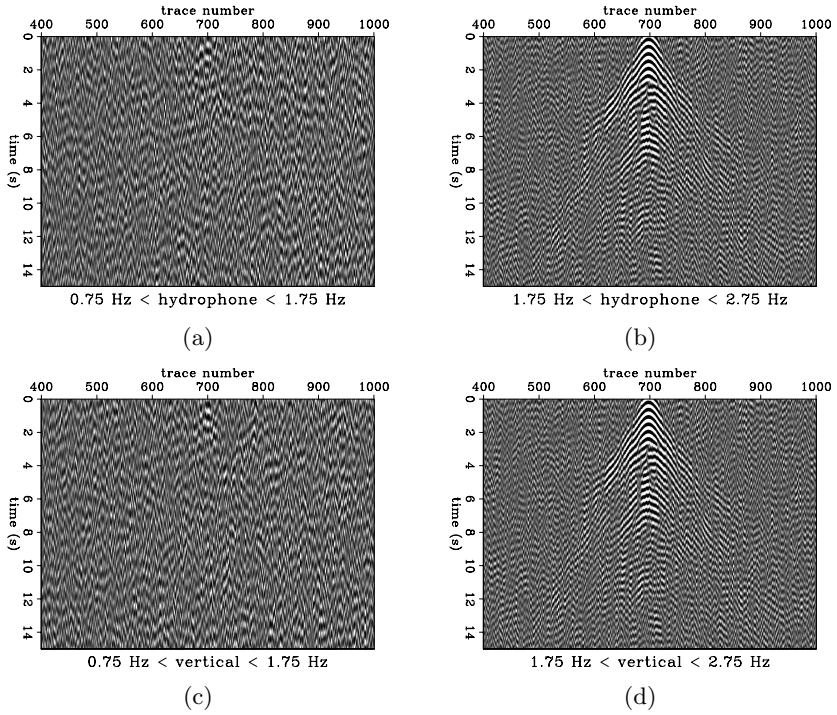


Figure B-3: a-b) Hydrophone common receiver gather filtered below and above 1.75 Hz respectively. c-d) Vertical component geophone common receiver gather filtered below and above 1.75 Hz respectively. [**ER**] sjoerd2/. gatherHlow,gatherHhigh,gatherVlow,gatherVhigh

SEG 2014 benchmark data

Ali Almomin, Xukai Shen, Carlo Fortini, Guillaume Barnier and Biondo Biondi

ABSTRACT

A synthetic bunchmark test data was created by Chevron for the SEG 2014 workshop. We first overview the data and the potential challenges that it presents for velocity estimation methods. Then, we process the data by reducing the noise and the sea bed surface-related multiples. Finally, we estimate the P-wave velocity model with a sequence of full waveform inversion methods that first analyze the early arrivals then the reflection data. The angle-domain common-image gathers of the final velocity model show significant improvement compared to the initial velocity model and indicate that most of the kinematics of the data were successfully estimated.

INTRODUCTION

Full waveform inversion (FWI)(Tarantola, 1984; Pratt, 1999) has become an increasingly popular technique to estimate the subsurface property from seismic data due to its high resolution and very accurate results. This high resolution is due to utilizing the information from both the forward-scattered and back-scattered wavefields in the data, i.e. both their kinematics and their dynamics. Moreover, the data misfit is computed in the data spaces. This direct computation of the errors results in a relatively simple relationship between the data residuals and the model updates. However, FWI has the disadvantage that its objective function is far from being smooth and convex which, in the absence of very low frequency, requires the starting model to be very close to the true model to avoid converging to a local minimum.

To test and compare different FWI algorithms, Chevron has recently released a benchmark synthetic dataset for a blind test as part of the SEG 2014 workshop. The data are modeled based on real geology and present several challenges that can prevent a conventional implementation of FWI from working properly.

We reduce the noise in the data by applying a bandpass filter followed by a multiples reduction algorithm that targets sea bed surface related multiples. To estimate the P-wave velocity model, we first run an early arrival full waveform inversion to estimate the shallow part of the model. Then, we run a low frequency tomographic full waveform inversion (TFWI) (Almomin and Biondi, 2014) on the reflection data to improve the background velocity at the deeper part of the data. Finally, we run a kinematics based FWI on the reflection data. The angle-domain common-image gathers of the final velocity model show significant improvement compared to the initial velocity model and indicate that most of the kinematics of the data were successfully estimated.

DATA OVERVIEW

The dataset was modeled using a 2D isotropic, elastic wave-equation with a free-surface at top and absorbing boundaries on the sides and the bottom. The source function used was provided as a farfield wholespace wavelet without ghosts effects with a sampling rate of 0.666666 ms. Figure 1 shows the wavelet provided and figure 2 shows the amplitude spectrum of the wavelet, which shows a majority of the signal between 2 Hz and 40 Hz with a dominant frequency of 10 Hz. A marine geometry was used with 1600 shots separated by 25 m and 321 hydrophone receivers separated by 25 m. The receivers start at the source location and reach a maximum of 8000 m. Both sources and receivers are at a depth of 15 m. The recorded data had a duration of 8 s with a sampling of 4 ms.

Figure 3 shows the first shot gather and figure 4 shows the zero-offset gather of the raw data, which show the source and receiver ghosts effects and the free-surface multiples. Moreover, noise was added to the data that mostly affects the very low and very high ends of the frequency spectrum. The noise can be seen in the amplitude spectrum of the first shot gather in figure 5 as well as the average amplitude spectrum of all shot gathers in figure 6, which was obtained by first computing the amplitude spectrum of each shot then averaging those spectra.

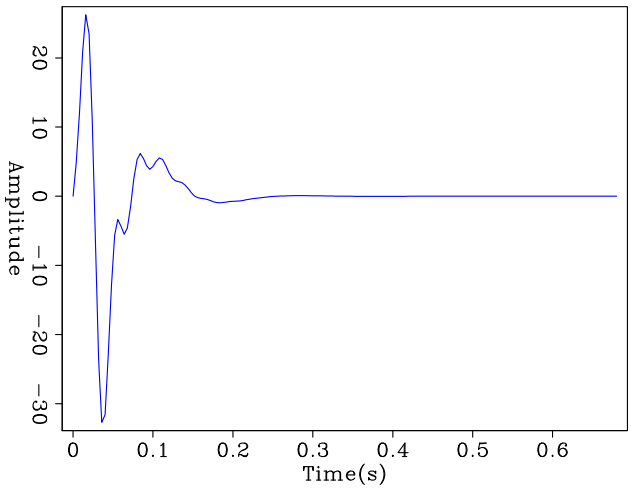


Figure 1: Source wavelet. [**ER**] ali2/. wavelet

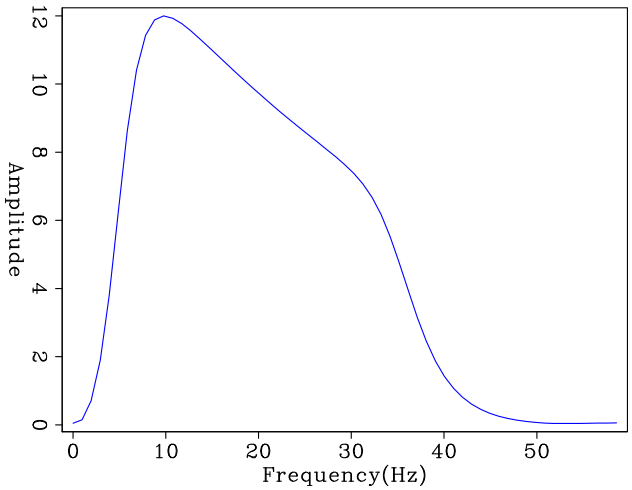


Figure 2: Amplitude spectrum of the source wavelet. [**ER**] ali2/. waveletspec

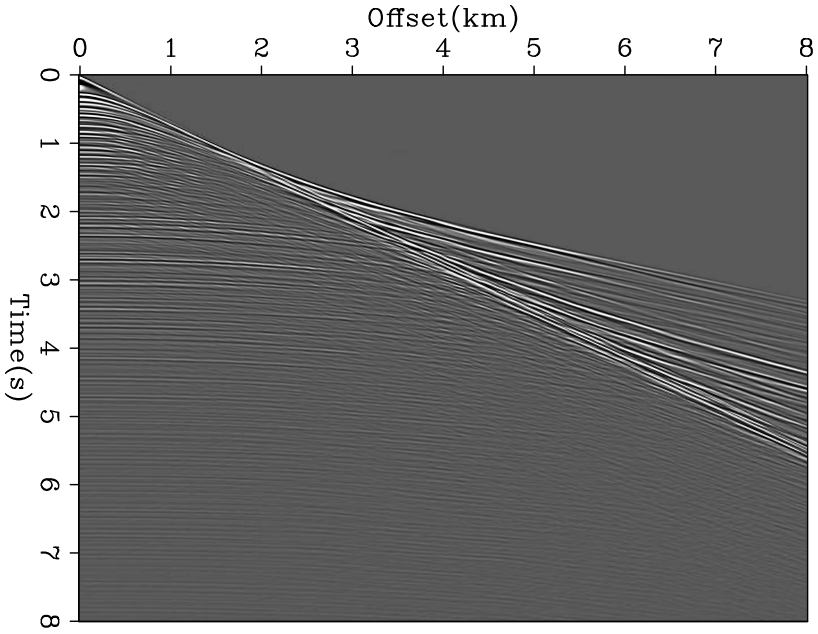


Figure 3: First shot gather of the raw data. [**ER**] ali2/. 1shot

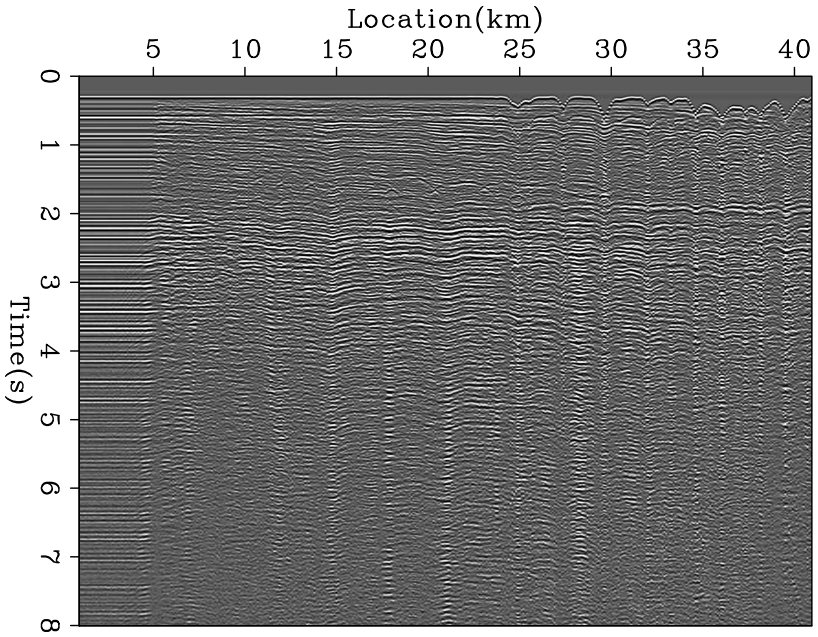


Figure 4: Zero offset gather of the raw data. [**ER**] ali2/. zostack

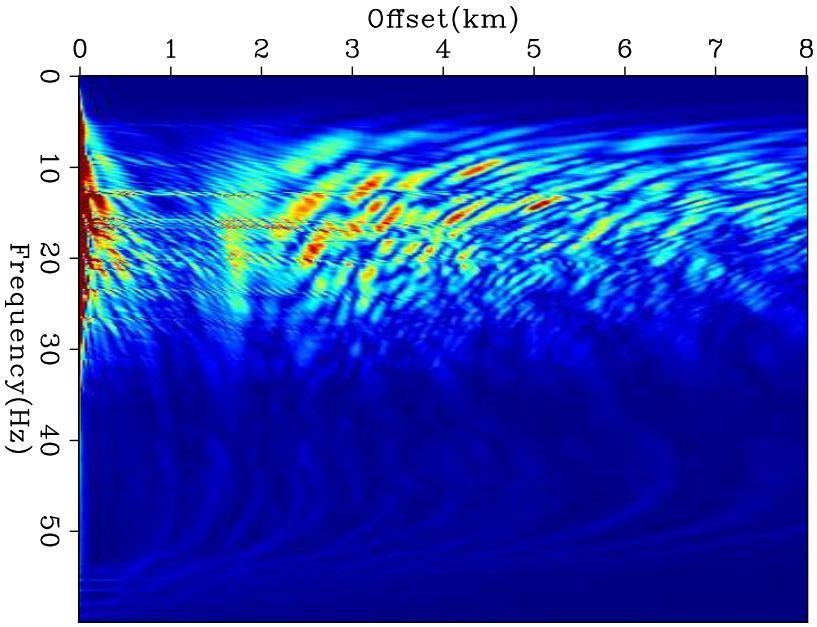


Figure 5: Amplitude spectrum of the first shot gather of the raw data. [**ER**] ali2/. 1shotspec

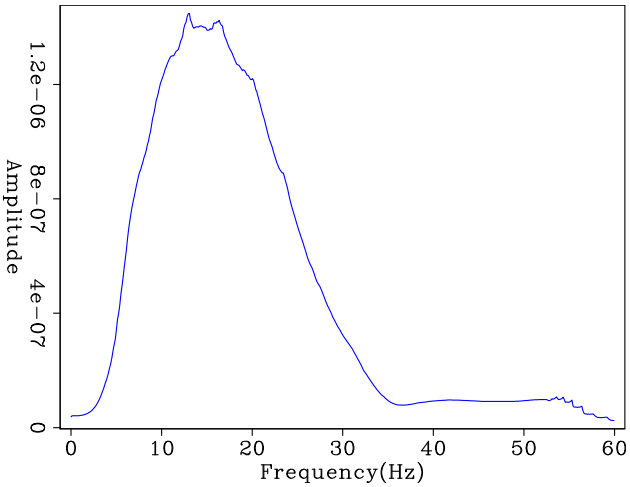


Figure 6: Average amplitude spectrum of all shot gathers of the raw data. [**ER**] ali2/. allshotspec

MULTIPLES REDUCTION

The original data was modeled with an elastic FD algorithm with a free-surface boundary condition in order to simulate the presence of surface-related multiples. The first layer of the model is constituted by water. The surface-related multiple events that underwent to, at least a bounce, at the depth of the seabed seem to be the most energetic ones.

In order to remove the multiples from the original data, we applied a model-based prediction algorithm that generates a model of the seabed surface-related multiples (noise model). We then adaptively subtracted the noise model from the original data with a 2D Least Squares adaptive subtraction filter.

The model-based algorithm we used is called MRKE (Multiple Removal by Kirchhoff Extrapolation). It needs as input an interpreted reflector and the velocity model between the acquisition surface and that reflector. By using a Kirchhoff wavefield extrapolation operator, the algorithm simulates the additional bounce of the primary events recorded in the data that occurred at the depth of the reflector. This way, the primaries contained in the data are transformed into multiples (and the multiples in higher order multiples). Since the same operation is applied on all the events in the original data, also the receiver-side peg-legs of the interpreted reflector are predicted. The source-side peg-legs can also be predicted by applying the same algorithm after a re-ordering of the input data in common-receiver gathers (CRGs).

Here are the steps we followed for the removal of the multiples from the original data:

1) Picking of the seabed interface from the given initial velocity model
2) MRKE on the original data
3) MRKE on the original data re-sorted in CRGs
4) Merge of the results of steps 2) and 3)
5) Adaptive subtraction of the predicted multiples from the original dataset

Figure 7 shows the first shot gather after removing the multiples. By comparing figure 7 to figure 3, we can see that several events, most notably around early time and far offset, have been removed. Figure 8 shows the amplitude spectrum of the shot gather shown in figure 7. Figure 9 shows the zero-offset gather with a t-power gain using an exponent value of one. Finally, figure 10 shows the average amplitude spectrum of all shot gathers, which was obtained by first computing the amplitude spectrum of each shot then averaging those spectra.

VELOCITY ANALYSIS

Initial velocity

A 1-D initial velocity model was provided with the data that contains the sea bottom and a water velocity of 1.51 km/s, as shown in figure 11. The left side of the model has a flat

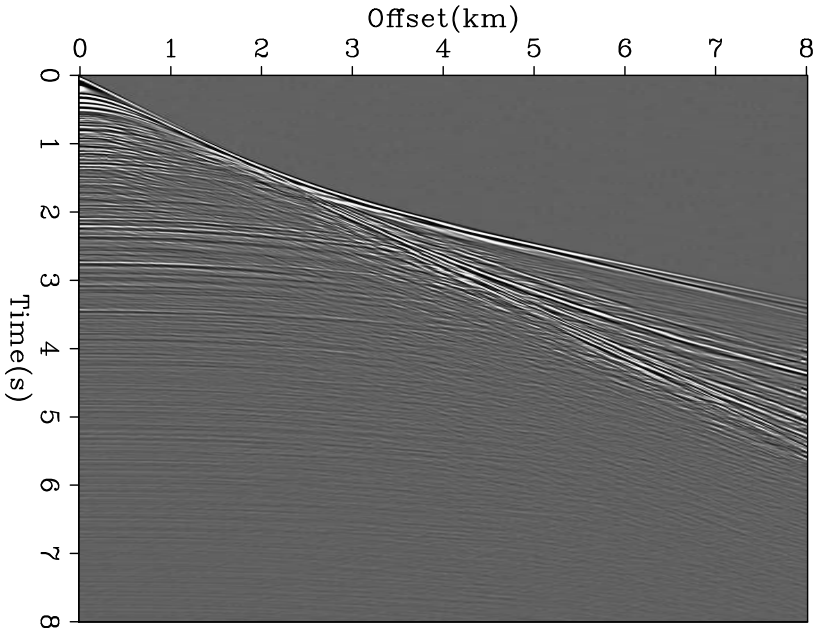


Figure 7: First shot gather of the de-multipled data. [**NR**] ali2/. 1shotdemult

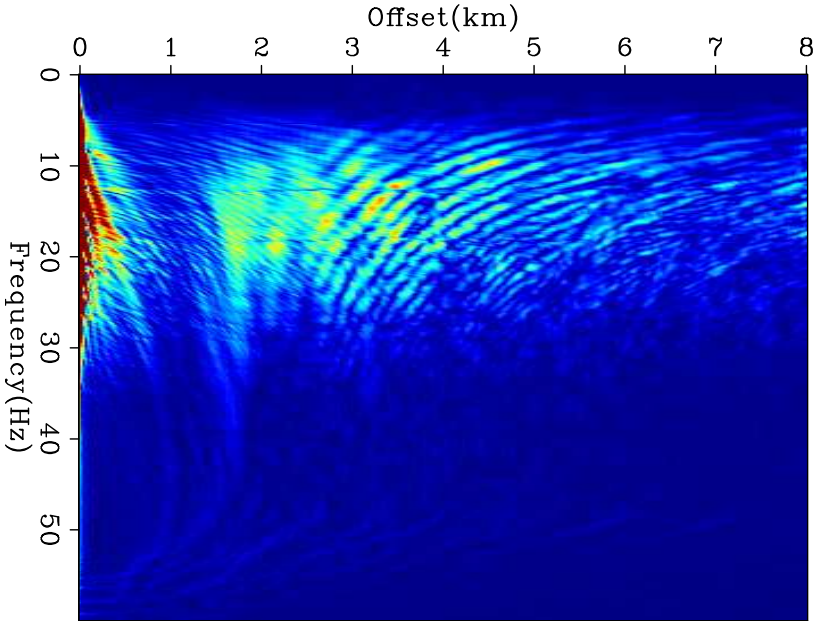


Figure 8: Amplitude spectrum of the first shot gather of the de-multipled data. [**NR**] ali2/. 1shotdemultspec

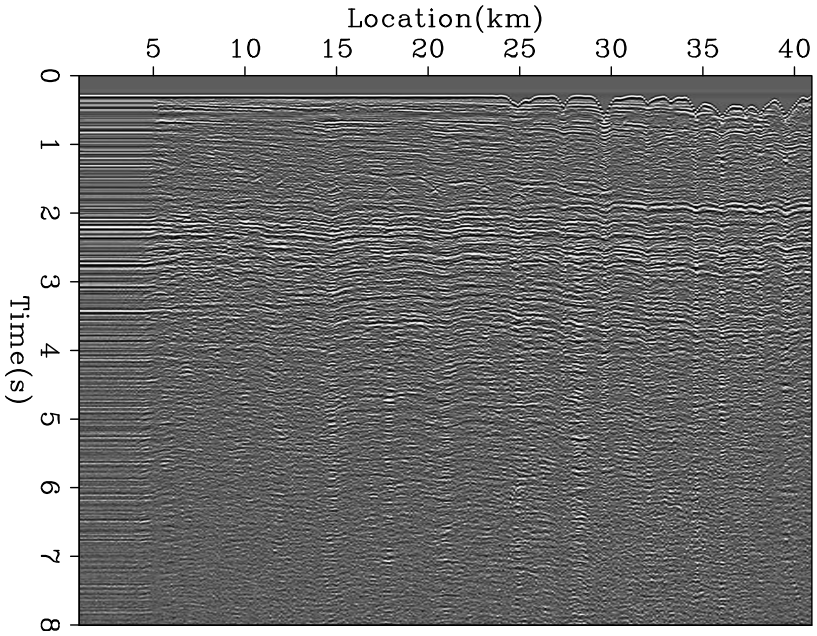


Figure 9: Zero offset gather of the de-multipled data. [**NR**] ali2/. zostackdemult

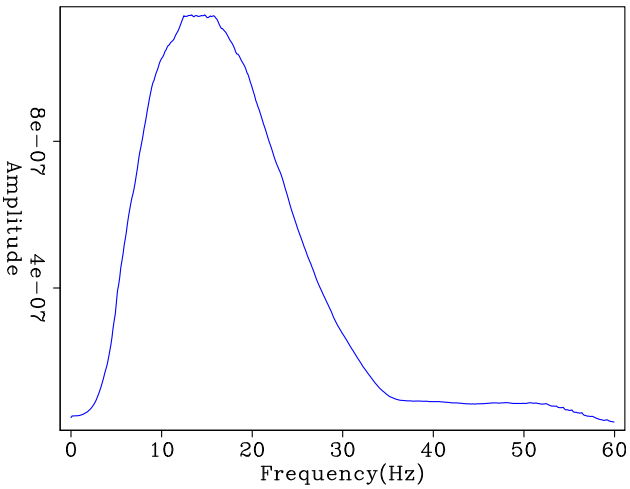


Figure 10: Average amplitude spectrum of all shot gathers of the de-multipled data. [**NR**] ali2/. allshotdemultspec

sea bottom whereas the right side has more complex topography. The first shot is located at 1 km with the receivers extending towards the right.

For all subsequent steps, we reduce the computational cost by subsampling the source function to a rate of 4 ms, matching the sampling of the recorded data, and the source spacing to a rate of 100 m spacing by keeping every fourth shot gather. We also applied a bandpass filter with parameters 2-4-35-40 Hz to reduce the low- and high-frequency noise. Moreover, we symmetrize around common-midpoint (CMP) locations using reciprocity in order to produce image and angle-gathers that are easier to interpret.

Figure 12 shows the migration image obtained using the initial velocity model and the demultipled data after muting the early arrivals. The events are very incoherent and defocused due to large velocity errors. Figure 13 shows the angle-domain common-image gathers (ADCIGs) using the initial velocity model. The ADCIGs have 13 equally spaced gathers starting at a location of 10 km and ending at 40 km with an opening angle range between -50 to +50 degrees. The gathers show strong curvature as we go towards the right side and the deep part of the model, indicating significant velocity errors in the model.

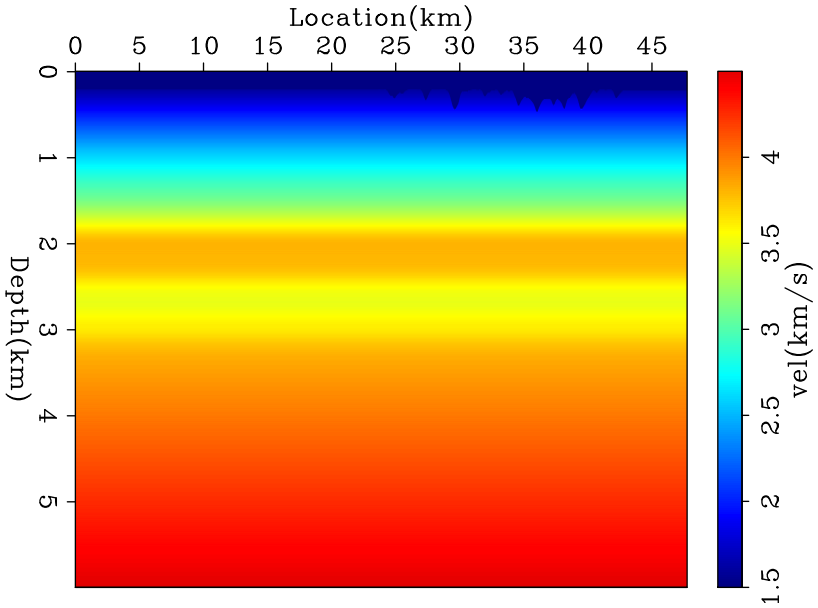


Figure 11: Initial velocity model. [**ER**] ali2/. v0

Early arrival FWI

The long-offset shot gathers provide an excellent opportunity for deriving the near-surface velocity model using early-arrival waveform inversion. The recorded data have offset up to 8 km, which allow us to estimate velocity down to about 2 km. The near-surface velocity estimation was performed using a workflow described in (Shen et al., 2012). Using first-breaks only, wave-equation traveltime inversion (Woodward, 1989; Luo and Schuster, 1991) was performed, using the preprocessed data with bandpass between 5 Hz and 10 Hz. After that, using early-arrivals with the same bandpass parameters as observed data, Kinematic Waveform Inversion (KWI)(Shen, 2014) was performed.

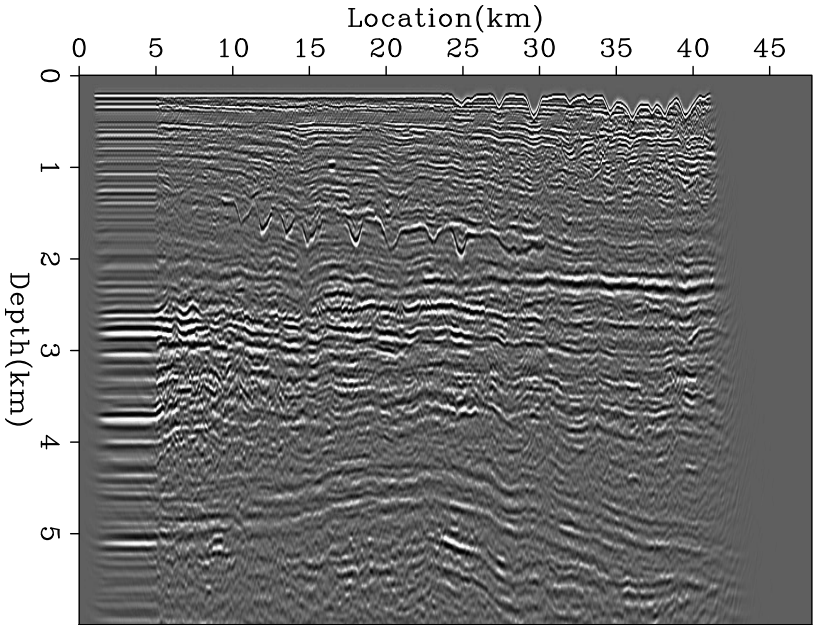


Figure 12: Migration image using the initial velocity model. [**NR**] ali2/. image0

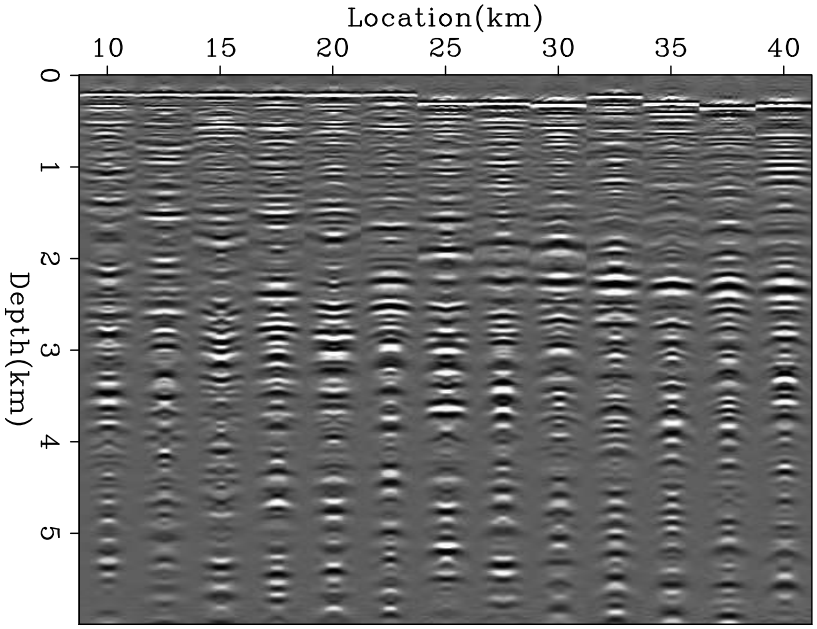


Figure 13: ADCIGs using the initial velocity model. [**NR**] ali2/. adcig0

The final result of the workflow in figure 14 shows significant improvement of the near-surface velocity model. Not only does the migrated image have much better lateral coherency (figure 15), but also does the angle gathers (figure 16) become much flatter

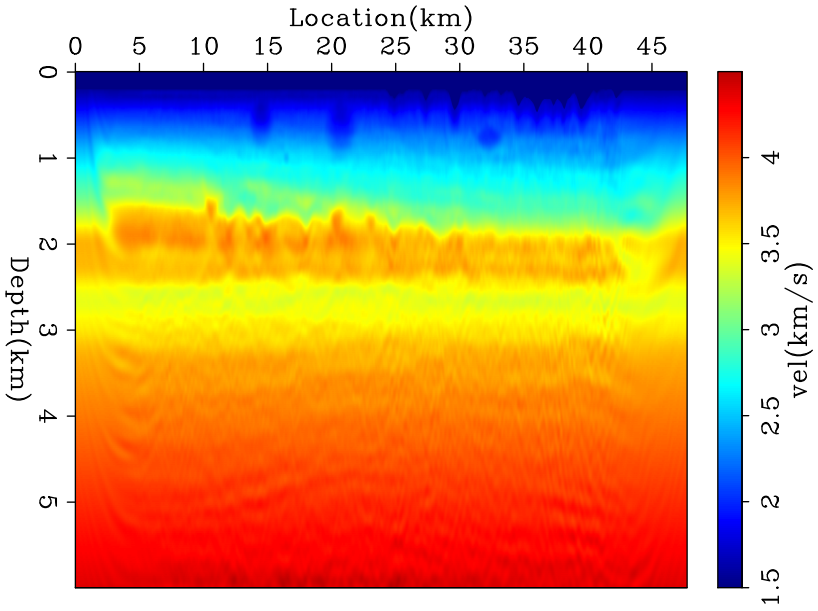


Figure 14: Early-arrival FWI velocity model. [**NR**] ali2/. v1

Low frequency TFWI

The next step is to extract the kinematic information from the reflection data. We follow the strategy suggested by Biondi and Almomin (2014) by doing a two-stage inversion. First, we ran TFWI on the low frequency part of the data (up to 10 Hz) with strong regularization. Second, we run FWI starting from the low-frequency TFWI results. The regularization used in TFWI can be equivalently applied as a preconditioning to the gradient. The preconditioner strongly smooths in the horizontal direction and slightly in the vertical direction. The preconditioning also helps avoiding overfitting the amplitude, which is undesirable in this case due to the fact that we use an acoustic wave-equation on data modeled with an elastic wave-equation.

The output of running low-frequency TFWI is shown in figure 17. There is a significant decrease in velocity between 2.5 km and 3.5 km depths and an increase in velocity below that. Also, a slow-velocity anomaly shows up around 28 km in the horizontal coordinate and 2.8 km depth. The migration image using the low-frequency TFWI is shown in figure 18, which shows a large improvement in reflectors continuity and amplitude compared to figure 15, indicating better overall focusing of the events. The improvement in velocity is also validated by flatness of reflectors as shown in figure 19.

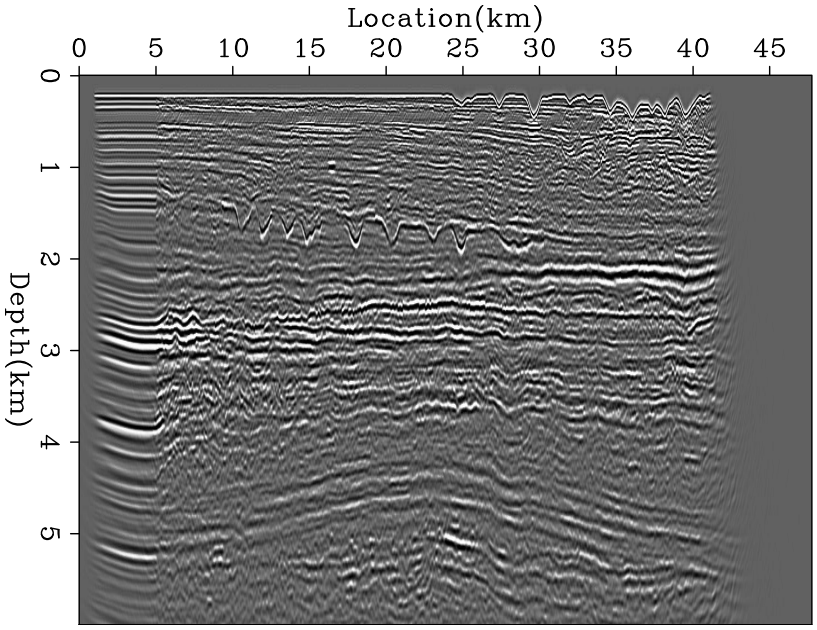


Figure 15: Migration image using the early-arrival FWI velocity model. [**NR**] ali2/. image1

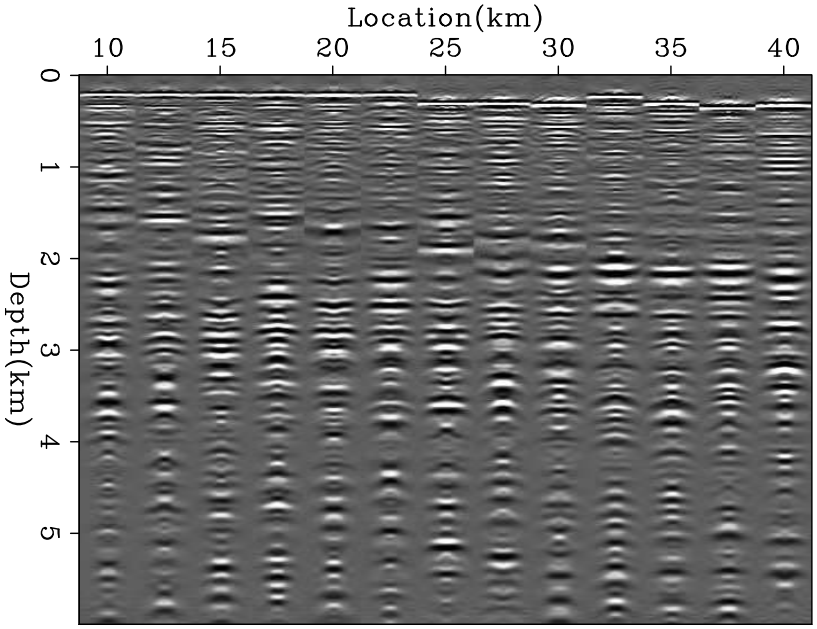


Figure 16: ADCIGs using the early-arrival FWI velocity model. [**NR**] ali2/. adcig1

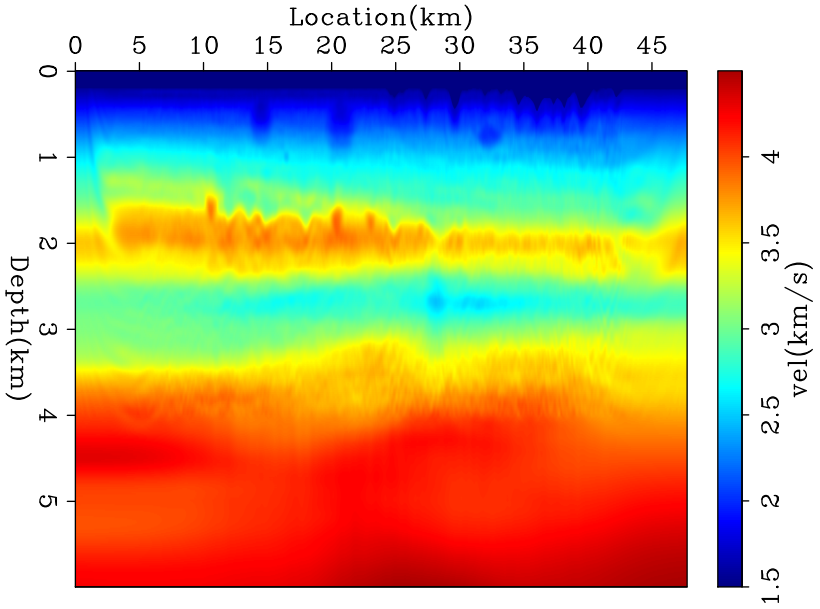


Figure 17: Low-frequency TFWI velocity model. [**NR**] ali2/. v2

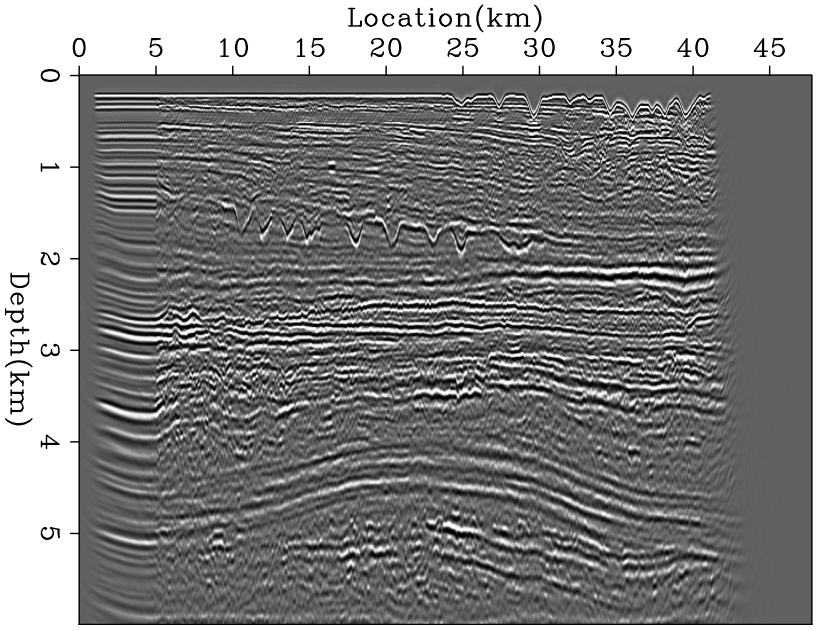


Figure 18: Migration image using the low-frequency TFWI velocity model. [**NR**] ali2/. image2

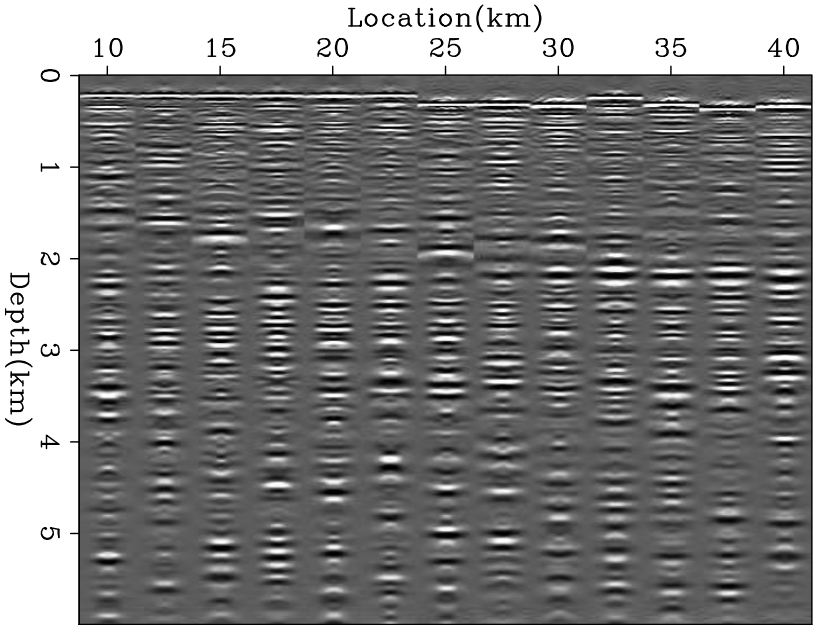


Figure 19: ADCIGs using the low-frequency TFWI velocity model. [**NR**] ali2/. adcig2

Reflection FWI

Next, waveform inversion was performed using the TFWI result as the starting model. Since early arrivals had been used to derive the near-surface velocity model, and the near-surface model almost did not change after TFWI, those early-arrivals were not used in the waveform inversion here. In other words, only reflections were used as the observed data. Consequently, the final result after waveform inversion (Figure 20) has a lot of details of the reflectors, especially after 2.5 km depth, where the previous early-arrival waveform inversion failed to update due to limited illumination. Thanks to the long-wavelength components updated from TFWI, the reflection waveform inversion was able to update the short wavelength components of the model. This is verified by better focusing of the migrated image (Figure 21) and better flattened angle gathers (Figure 22).

CONCLUSIONS

Using the Chevron SEG 2014 synthetic dataset, we estimated the velocity model in a blind FWI test. We first processed the data to enhance the signal-to-noise ratio by bandpassing the useful part of the spectrum and reducing the surface related multiples using the MRKE algorithm. Then, we estimated the velocity in three stages: early-arrival FWI, low-frequency TFWI, and reflection FWI. The early-arrival FWI stage improved the shallow part, the low-frequency TFWI improved the kinematics of the deep part, and the reflection FWI added the reflection information to the model. The migration images as well as the ADCIGs indicate that the final velocity model has successfully estimated the majority of the kinematics in the data. However, we did not fully fit the amplitude due to the difference in physics between the observed data and the modeling operator.

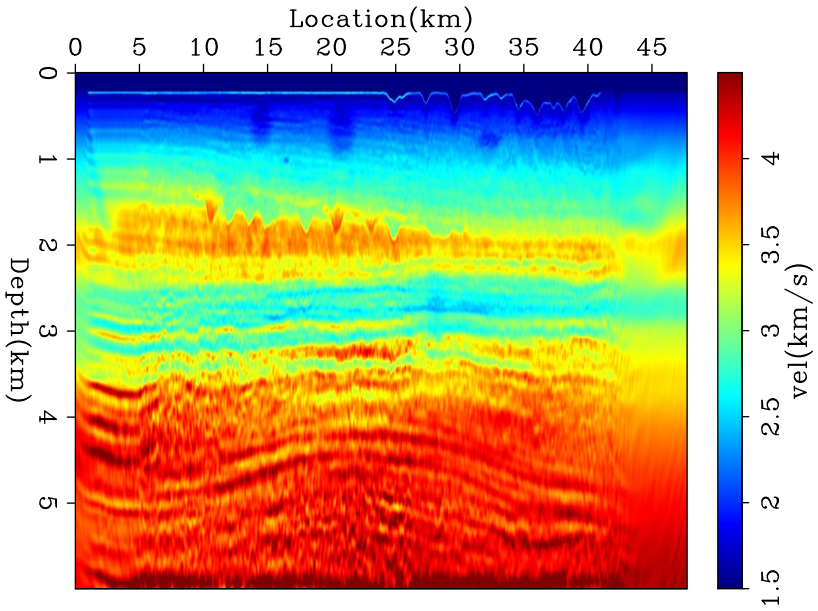


Figure 20: Reflection FWI velocity model. [**NR**] ali2/. v3

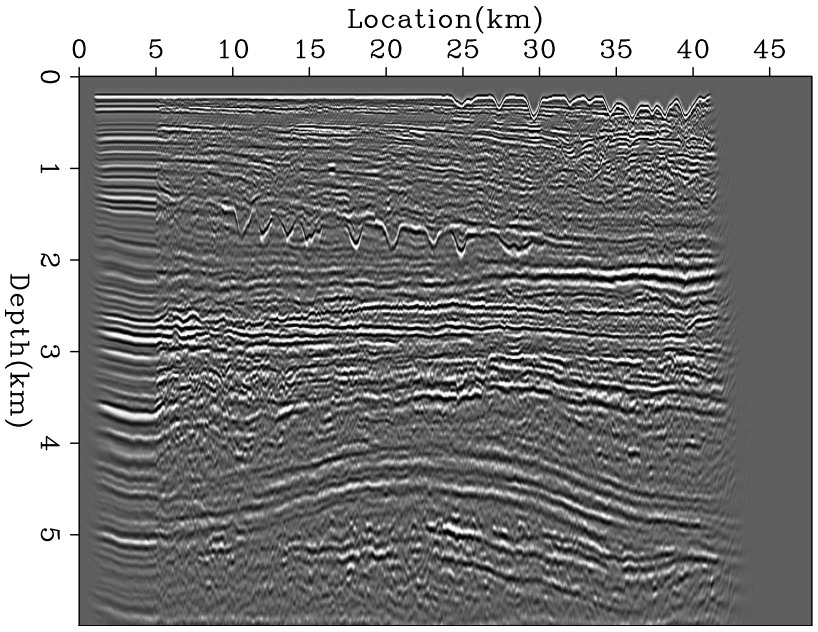


Figure 21: Migration image using the reflection FWI velocity model. [**NR**] ali2/. image3

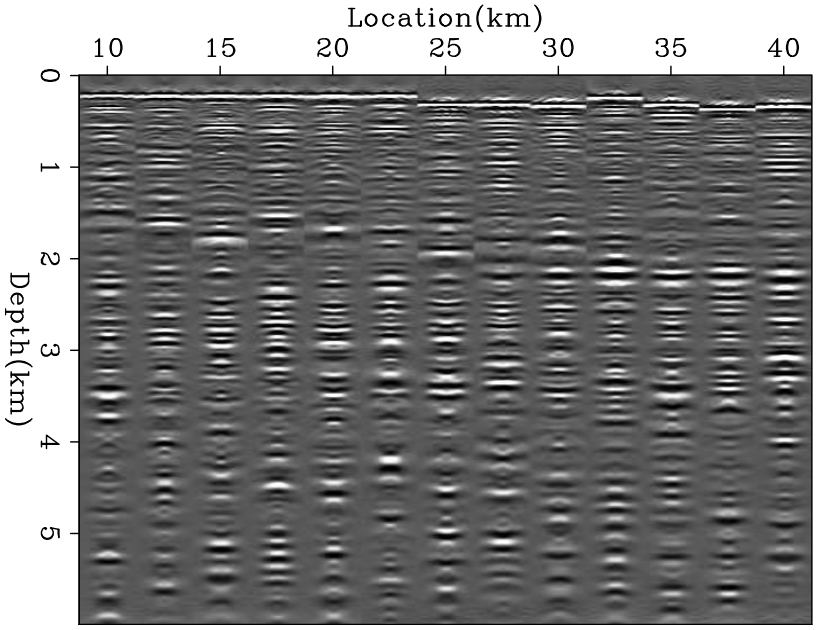


Figure 22: ADCIGs using the reflection FWI velocity model. [**NR**] ali2/. adcig3

ACKNOWLEDGEMENT

We would like to thank Chevron for providing the dataset and permission to publish the results.

REFERENCES

Almomin, A. and B. Biondi, 2014, Preconditioned tomographic full waveform inversion by wavelength continuation : SEP-Report, **152**, 11–18.

Biondi, B. and A. Almomin, 2014, Efficient and robust waveform-inversion workflow: Tomographic FWI followed by FWI : SEP-Report, **152**, 1–10.

Luo, Y. and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**, 645–653.

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model: Geophysics, **64**, 888–901.

Shen, X., 2014, Early-arrival waveform inversion for near-surface velocity estimation: PhD thesis, Stanford University.

Shen, X., T. Tonellot, Y. Luo, T. Keho, and R. Ley, 2012, A new waveform inversion workflow: Application to near-surface velocity estimation in saudi arabia: SEG Abstracts.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

Woodward, M., 1989, Wave equation tomography: PhD thesis, Stanford University.

Irregularly-spaced, non-stationary signals

Jon Claerbout

ABSTRACT

Of three methods to deal with nonstationary signals the most appealing is to interpolate filters from a coarse mesh. While operators conveniently carry to a regular mesh scattered data *values*, scattered data *signals* invite techniques more akin to matrix inversion.

INTRODUCTION

My new book, to be my final book, GIEE (Geophysical Image Estimation by Example)[1] scarcely touched two important areas of application: (1) Dealing with nonstationarity by interpolating filters from a coarse mesh, and (2) Carrying irregularly-spaced geophysical data *signals* as opposed to *values* to a regular mesh. Here I provide some background for these two areas. Ultimately, both problems should be handled simultaneously.

NONSTATIONARY OPERATORS

Nonstationary data are those with spectra changing in time or space. The most common form of nonstationarity is waves changing their direction with time and space. Nonstationary data usually calls for nonstationary operators. We need those to accelerate solutions, to fill in data gaps, and to transform residuals to whiteness (IID).

Time-variable 1-D filter

My first go at nonstationarity was a time-variable PEF. Unfortunately, at the present state of computer hardware, the method is not suitable for multidimensional data. This method did work well in one dimension. Figure 1 shows synthetic data with time variable deconvolution. (More details are in the document labeled "Unfinished" at my website.)

The method is simple. Every point on the signal has its own filter. Because each data point has a multi-point filter, the PEF-design regression is severely underdetermined; but a workable regularization is forcing filters to change slowly. I minimized the gradient with time of the filter coefficients.

As we hope for deconvolution, events in Figure 1 are soon compressed to impulses. The compression is remarkably good, even though each event has a different spectrum. What is especially pleasing is that satisfactory results are obtained after truly small numbers of iterations (roughly three). The example is for two free filter coefficients $(1, a_1, a_2)$ per output point.

[1]Not yet available on the retail market. Available from the manufacturer at lulu.com.

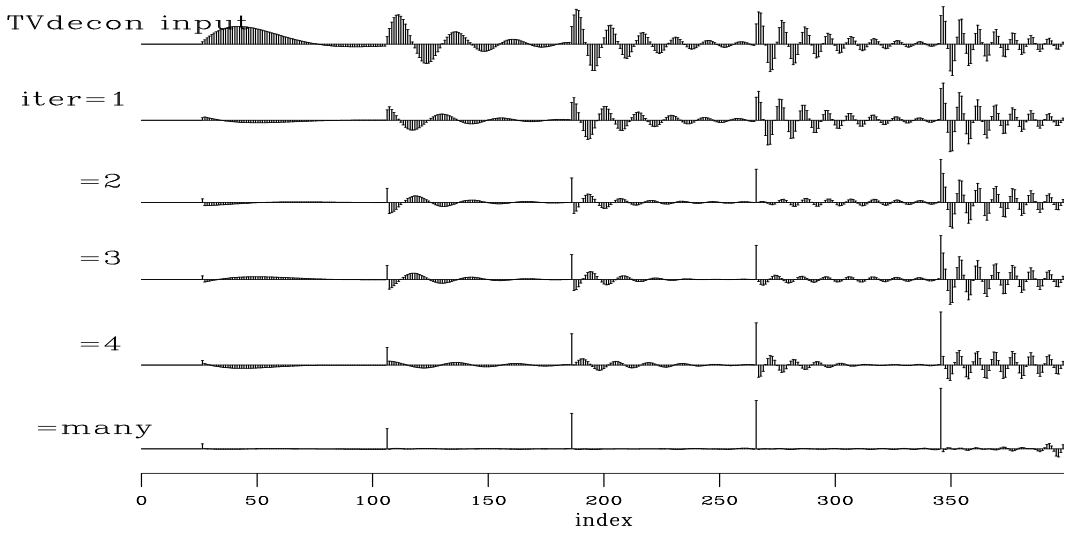


Figure 1: Time variable deconvolution with two free filter coefficients and a gap of 6. [**ER**] jon/. tvdecon90

Dip spectra commonly vary in time and space. In multidimensional spaces, we primarily struggle for machine memory. There, needing a filter array for each data point is abhorrent.

Patching

My second go at nonstationarity was patching. A big block of data is chopped into overlapping little blocks. The adjoint operation merges the little blocks back into a big block. The inverse patching operator is easily found by passing a big plane full of ones through the operator and back. What emerges will measure the overlap, i.e., find a bin count for a divisor to convert the adjoint to an inverse. Weighting functions of space may also be introduced and the inverse likewise calculated. Patching would appear to be well suited to modern parallel computer architectures.

Patches need not be equal in size nor be rectangular. Reflection seismologists immediately recognize the need for wedge-shaped patches in the space of time and source-receiver offset.

This method does work, but there are drawbacks. A big drawback is the many parameters required to specify patch sizes and overlaps. When PEFs are designed in blocks, then care must be taken to use internal filtering and attend to the fact that output lengths are shorter than input lengths. You live in fear that patch boundaries may be visible in your output. The many parameters increase the likelihood of miscommunication between the coder and the user. The many parameters also require effort and experience to tune.

Store the filter on a coarser mesh.

The first coarse-mesh filter idea is to keep the filter constant over a range of values in time and space. Such a filter would be easily stored on a coarser mesh, so the memory devoted

to filters could be significantly less than the data. But, this idea evokes fear the outputs may show the blocky boundaries.

Bob Clapp (who has exercised nonstationary filtering in large-scale environments) suggests we should linearly interpolate filters from the coarser mesh. It can become costly, but economics are hard to figure in this age of rapidly changing computer architectures. Whether or not and how the coarse-mesh-filter idea is integrated with the helix transform is a topic that to my knowledge has not yet been attacked. The challenge for the analyst/-coder is to produce filters interpolated from a grid in an environment that can be widely shared among many applications and with many people.

MOVING IRREGULARLY-SPACED SIGNALS TO A REGULAR MESH

Although we now have much experience taking data to a regular mesh bringing irregularly-spaced *scalars,* for irregularly-spaced *signals* we must do the problem at each time point, repeatedly solving it for each. There are, however, thousands of time points on a seismogram. Yikes! We need some way of accumulating and reusing knowledge. It seems we need something like an inverse matrix, but, that is exactly what GIEE has avoided, the reason being to avoid hopelessly large memory requirements.

We have accomplished much by using operators (function pairs) instead of matrices (data structures). We will soon see here that irregularly-spaced signals suggest a need for switching from operators to sparse matrices. Often, we move irregularly-spaced data to a regular mesh. We want the regular mesh for data viewing, filtering, correlation, and Fourier transformation. When we have data signals instead of simply data values, it seems we must repeatedly use the same iterative program at each point in time. But, sparse matrices might be vastly faster. To see why, represent a large collection of least-squares regressions (N time points), each with the same time-independent operator $\mathbf{F}^T\mathbf{F}$.

$$\mathbf{F}^T\mathbf{F}\,[\,\mathbf{m}_1\mathbf{m}_2\mathbf{m}_3\cdots\mathbf{m}_N] \quad = \quad \mathbf{F}^T\,[\mathbf{d}_1\mathbf{d}_2\mathbf{d}_3\cdots\mathbf{d}_N] \quad = \quad [\mathbf{b}_1\mathbf{b}_2\mathbf{b}_3\cdots\mathbf{b}_N] \tag{1}$$

Instead of iterating the same operator at each time, efficiency might be gained by approximating $(\mathbf{F}^T\mathbf{F})^{-1}$. How big is $\mathbf{F}$? We often deal with seismogram numbers from a hundred to a hundred thousand. Scalar data sets in GIEE range in size from a quarter to a half million. Model spaces there are typically larger because of zero padding. Examples there are solvable in a few minutes in desktop computers using operators. Industrial settings have comparable numbers of signals as GIEE has values. Luckily, industrial and some academic settings have clusters of computer cores, today numbering hundreds to tens of thousands.

Outside suggestions

I asked Michael Saunders for an approach using operators. He made two suggestions: First, scan the research literature, and Second, consider instead sparse matrices, in particular a technique known as sparse QR.[2]

[2]Michael Saunders recommends http://www.cise.ufl.edu/research/sparse/SPQR/ a sparse QR method and code by Tim Davis.

Think of a sparse matrix $F_{i,j}$ as a list of three columns and K rows. Each row contains (matrix element, i value, j value). Observe how to multiply a sparse matrix times a vector, say $\mathbf{d} = \mathbf{Fm}$ or $d_i = \sum_j F_{i,j} m_j$.

```
do k=1,K
     data(i(k)) += matrix(k) * model(j(k))
```

A singular contribution of GIEE is multidimensional PEFs. Unlike gradient and Laplacian, PEFs are easily invertible, offering solution via a preconditioner $\mathbf{p}$ in which $\mathbf{m} = \mathbf{A}^{-1}\mathbf{p}$. In the simplest case $\mathbf{A}^{-1}$ would be leaky integration, trivially implemented with recursion. Recursion allows an easy solution to these huge problems. In the QR algorithm, recursion appears as the triangular matrix $\mathbf{R}$, in the name QR.

My preliminary ideas for how to do it

In practice the model mesh may always be dense enough that linear interpolation is adequate. We start from this assumption. As warm up, think about only one data signal in 2-D model space. On a first iteration, adjoint interpolation brings the data signal to its neighboring four mesh locations. A small number of iterations brings it to the surrounding neighborhood. When we need not fill a large region, not many iterations are required. In practice we push all data signals to the mesh at the same time. However, each time level requires us to solve an identical iterative problem. As there are typically thousands of time points, those iterations get tiresome! Let us solve this problem at each of about 40 time levels. Then let us see how we might use these results to more quickly obtain mesh values at the remaining thousands of time levels.

Limiting calculation to the 40 time levels, consider each mesh signal separately. Correlate the mesh signal with all the data signals. Select data signals with the strongest correlation. Using only those data signals, find the coefficients defining the best linear combination of data signals. Use these coefficients to define this mesh signal for all other times. The idea of using only data strongly correlated to the mesh signal could be made more sophisticated, and perhaps better. Limiting to 40 time points, using all data signals to fit the mesh signal we could jettison data signals from the fitting by applying some ℓ_1 penalty to the fitting coefficients.

Alternate view

I would be more satisfied with this algorithm if instead of 40 time levels, it dealt with 40 time lags. But I don't know how to put that together.

CONCLUSION

Well my friends, we have come a long way, and made much progress. I have grown old, so it may be up to you to flesh out the theory, write the code, and produce the examples, thereby uncovering the pitfalls. I'll try to interest some young person to take on these projects, but don't hold your breath.

Iterative migration using sparseness constraints with the hyperbolic penalty function: Application on two 2-D field datasets

Mandy Wong and Antoine Guitton

ABSTRACT

Sparse reflectivities are obtained with iterative migration thanks to the hyperbolic penalty function for both data fitting and model styling goals. Sparseness is achieved by letting parts of the model treated in a "ℓ^1-norm" sense. Comparing with a more classical least-squares approach without regularization, the sparse-reflectivity images have less artifacts and better defined reflectors. However, these reflectors are often less continuous and the parameterization of the hyperbolic penalty function remains cumbersome. The main advantages of the hyperbolic penalty function, as opposed to other "ℓ^1-type" norms, is that it is convex, can behave like the ℓ^1 or ℓ^2 norms as needed, and can be minimized very efficiently with a fast non-linear conjugate direction method.

INTRODUCTION

This manuscript presents the follow up of our previous work (Wong and Guitton, 2014) on applying the hyperbolic penalty function to the iterative migration problem. Iterative migration poses the imaging problem as an inversion problem. Instead of applying the migration operator, which is regarded as the adjoint to the forward modeling operator, iterative migration aims to apply the inverse of the forward modeling operator to recover the reflectivity series of the subsurface. Due to the size of the problem, iterative migration is usually performed iteratively using gradient based techniques. Many studies have shown that the least-squares iterative migration image has fewer migration artifacts, better relative amplitude information, and higher resolution than the corresponding migration image (Clapp, 2005; Valenciano, 2008; Wong et al., 2012). In Wong and Guitton (2014) and for iterative migration, we used the hyperbolic penalty function (HPF) in the regularization term to recover a sharp and sparse reflectivity model in a 1D synthetic example. In addition, we have shown that when data are contaminated with non-Gaussian noise (e.g, noise bursts, spikes), the HPF for the data-fitting part is more robust in recovering the reflectivity series. The HPF with the correct parameterization for measuring data and model misfits is often a better choice than the standard ℓ^2 norm because it varies smoothly between the ℓ^2 norm for small residuals and ℓ^1 norm for high residuals. Therefore, the HPF can be made robust to outliers present in the data and can help build sparse models. Being convex, a fast solver based on a non-linear conjugate direction method can be used to minimize the HPF efficiently (Claerbout, 2014).

We apply iterative imaging on two 2D field data examples using a HPF solver to get sparse reflectivity series and a ℓ^2 solver for our more traditional least-squares migration

results. Note that both solvers use the previous search direction and the current gradient to find the optimal search path (Claerbout, 2014). The first dataset covers the region of a layered earth. The second dataset is located in a region with a salt structure. Inversion results with both datasets show that the iterative migration results with the HPF in the regularization term are sparse, with less migration artifacts and better signal content than their least-squares migration counterparts. These results rely heavily on our ability to select a judicious set of parameters controlling the ℓ^2/ℓ^1 behavior of the HPF, as well as the strength of the regularization term.

In this paper, we begin by introducing the HPF and the objective function we are minimizing. Next, we discuss the objective functions used in this study. Then we show with 2D field data examples how the iterative migration performs with the ℓ^2 and HPF solvers.

THEORY

We now present the hyperbolic penalty function and its main properties, as well as the objective functions used in this study.

Hyperbolic Penalty Function

The HPF (Claerbout, 2014) is a convex penalty function that varies smoothly from ℓ^2 to ℓ^1. Equation (1) below presents the HPF, $H(r)$, and its derivatives:

$$\begin{aligned} H(r) &= \sqrt{1+(\tfrac{r}{g})^2}-1, \\ H'(r) &= \tfrac{r}{g\sqrt{1+(\frac{r}{g})^2}}, \end{aligned} \tag{1}$$

where g is a constant that scales the residual r. The first derivative, $H'(r)$, behaves like the first derivative of the ℓ^2 norm at small $|\frac{r}{g}|$. At large $|\frac{r}{g}|$, it behaves as the first derivative of the ℓ^1 norm. In practice, the factor g is often taken to be the value of some percentile of residual magnitudes. Its value determines what part of the residual is treated as ℓ^1 and what part is treated as ℓ^2. The HPF is minimized with a non-linear conjugate direction solver and is described in details in Claerbout (2014), Chapter 6.

Objective Function

In iterative migration, we expect the reflectivity model to be made up of a series of sharp and sparse signals. The basic objective function using the ℓ^2 norm is:

$$S_2(\mathbf{m}) = \|\mathbf{W}\left(\mathbf{L}\mathbf{m}-\mathbf{d}\right)\|_2^2, \tag{2}$$

where $\mathbf{L}$ represents the Born modeling operator acting onto the reflectivity model $\mathbf{m}$, and $\mathbf{d}$ is the observed data. $\mathbf{W}$ is a data weighting function.

To promote the sparseness in the reflectivity model $\mathbf{m}$, we include a model-styling goal using the hyperbolic penalty function as measure of fitness. The desired objective function

$S_{HPF}(\mathbf{m})$ becomes:

$$S_{HPF}(\mathbf{m}) = H_d(\mathbf{W}\,(\mathbf{Lm} - \mathbf{d})) + \epsilon H_m(\mathbf{m}), \tag{3}$$

where H_d and H_m are HPFs for the data-fitting and model-styling goals, respectively. Three constants need to be chosen in equation (3): two thresholds values g_d and g_m that regulate the transitioning behavior between ℓ^2 and ℓ^1 in the HPF, and ϵ that tunes the relative strength of the regularization with respect to the data-fitting goal.

After inspecting the two datasets, we concluded that no large spikes or bursts are present in the data. Therefore, we selected a large value for g_d such that the data-fitting function in equation 3 is effectively measured by the ℓ^2 norm. For simplicity, we will refer the inversion using equation 3 as the HPF iterative migration.

FIELD DATA EXAMPLES

The two 2-D datasets selected in this study were extracted from a 3-D survey carried out in the Gulf of Mexico. The maximum offset is about 4 km with a shallow water bottom at 500 m. This region has significant anisotropy (Li et al., 2014). Since we are using isotropic modeling, and considering that other physical processes are ignored (e.g., attenuation), we don't expect to fit all the information contained in the data.

We corrected for the 3D spherical spreading effects so that the data can better match the amplitude of the 2D modeling and migration operators. We apply the iterative migration in two regions: a sedimentary environment, and a region with a salt body. The inversion of these traces with and without HPF regularization, are presented here.

Sediment velocity model

Figure 1a shows a shot gather from the ExxonMobil data in the sediment area. This area contains mostly horizontally stratified layers with strong amplitudes recorded from shallower reflections. As a result, we apply a diagonal data weighting (Figure 1b) in the iterative migration to emphasize the reflection energy from the deeper part of the model. Also, the data weighting excludes the direct arrival and head waves.

Figure 2a shows the migration velocity model. The velocity values increase from 1500 m/s to 4000 m/s in depth. There is a fault cutting diagonally across the model starting from a depth of 2000 m. Figure 2b shows the reverse-time migration (RTM) image. Figure 3a shows the inverted image using least-squares reverse time migration (LSRTM) after 12 iterations of conjugate gradient algorithm. Equation 2 shows the objective function for LSRTM. When compared to the RTM image, we clearly see the increased bandwidth of the inverted image (deconvolution effect). Some of the weaker reflectors become more apparent over iterations as shown in the annotation. However, the inversion tends to boost migration artifacts as well, thus increasing the overall noise level in the image.

Figure 3b shows the inverted image at iteration 20 using the hyperbolic penalty function (Equation 3) for $\epsilon = 0.00125$ and $g_m = 12.5$. The model-styling term in the objective function has enforced sparsity in the reflectivity solution. The addition of the sparseness

constraints helps attenuate the artifacts in the shallow region that we could see in the least-squares migration result. However, the sparseness constraint tends to make some reflectors less continuous.

Figure 4a shows one shot gather from the input data. After 1 iteration, the data residual from LSRTM (Figure 4b) and HPF iterative migration (Figure 4b) drops significantly. At iteration 12, the least-squares migration data residual (Figure 4c) drops slightly with the stronger reflection energy being less coherent. Figure 4f shows the data residual from HPF iterative migration at iteration 20. Due to the presence of the model styling goal, the data fitting is a little bit less than the one in Figure 4c.

Figure 6 shows the normalized power spectra in the z-direction for the RTM, LSRTM, and HPF iterative migration images. Since the amplitude scale between migration and inversion images are different. We normalized the amplitude of the power spectra by scaling the maximum value to 1. We then converted the normalized amplitude into decibels. Notice that, with iterative inversion, the LSRTM and HPF spectrum becomes more white, especially in the higher and lower frequency range. The LSRTM spectrum shows a higher frequency content than the HPF one probably due to the increased noise content. In the HPF inversion, the high frequency noise is suppressed. The relative amplitude in the range of $kz = 0.004 - 0.012$ is higher than that from LSRTM.

Choosing epsilon and threshold values

Figure 5 shows 9 inverted images from the HPF iterative migration using different ϵ and g_m values for the sediment case. As expected, by increasing g_m for a constant ϵ the images become less sparse because more parts of the images are treated in a "ℓ^2 norm" sense and less in a "ℓ^1 norm" sense. Similarly, increasing ϵ for a constant g_m, the images get more sparse due to the increased influence of the regularization term in the objective function (equation 3). Our selection of both parameters as our best result is based on two criteria: data fit comparable to the LSRTM result and sparseness content. The later can be quite subjective and left to the interpreter's opinion. A more systematic procedure would be preferable with larger 3-D datasets.

Sediment velocity model with salt

Now, we study a 2D line that cuts across a salt body. Figure 7a shows a shot gather. There are strong and fast refraction events that come from the top of salt layer. Similar to the sediment case, a diagonal data weighting (Figure 7b) is used in the iterative migration. Figure 8a shows the migration velocity model. The velocity value in the sediment layer increases gradually with depth. The salt body is shown in red with a velocity value close to 5000 m/s. Figure 8b shows the RTM image in this region. There are low-frequency artifacts above the salt body due to the application of the imaging condition to waves, in the receiver and shot wavefields, traveling in the same directions. The shallower reflectors are not well imaged due to the lack of illumination. In addition, there is a large amount of noise (coherent or spurious) within the salt body. These artifacts are better suppressed in the LSRTM image (Figure 9a) after 20 iterations. However, the shallow region remains noisy. With HPF iterative migration (Figure 9b), the above mentioned artifacts are better

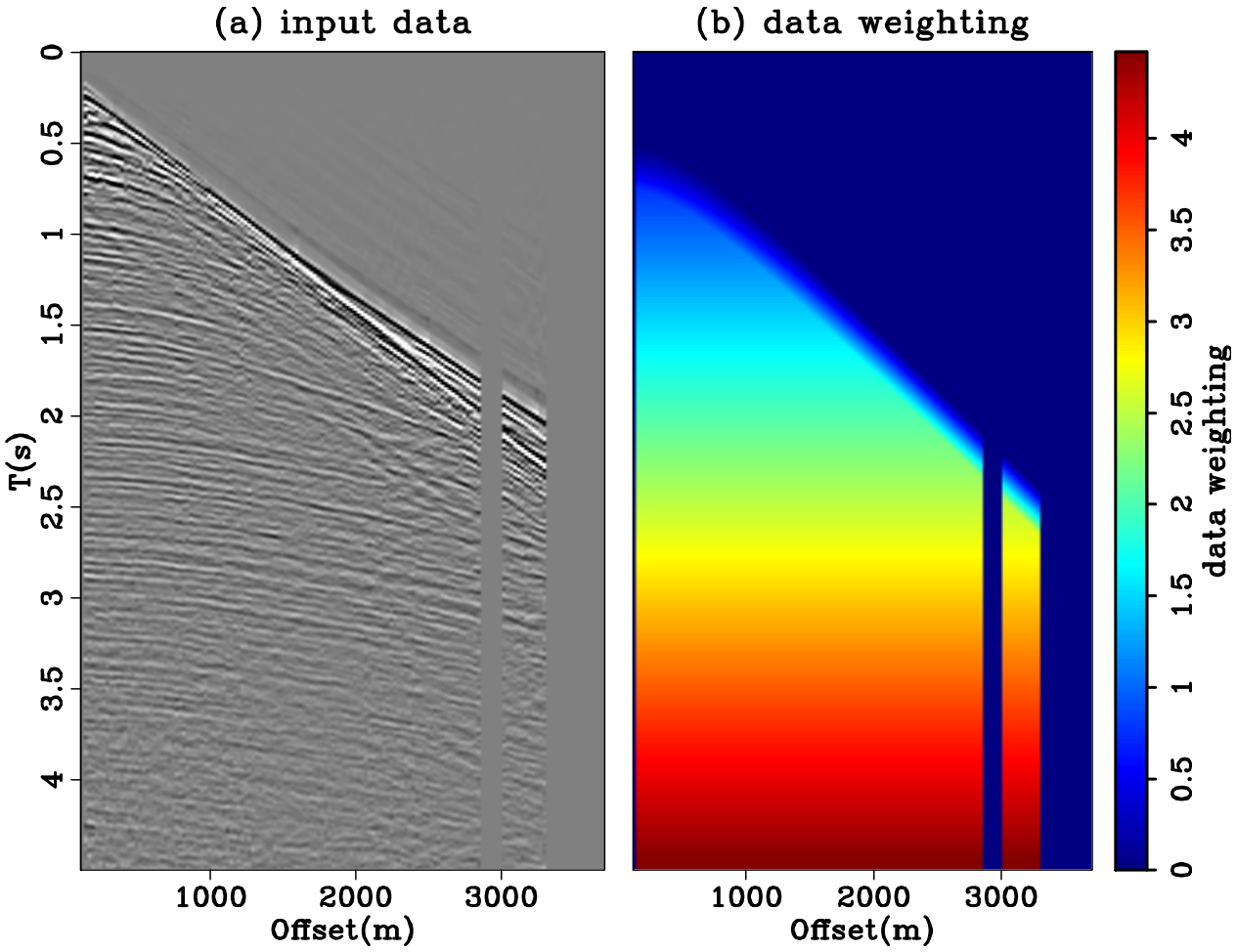


Figure 1: Input data for iterative migration in the sedimentary region. (a) is a shot gather. Notice the amplitude of the earlier reflection is much stronger than the deeper reflection. Therefore, (b) shows the data-weighting function used in the inversion to boost up the energy from the deeper part of the subsurface. The weighting function is also designed to exclude the direct arrival and head waves from the data-fitting part. [**ER**] mandy1/. fig1-inputdata

suppressed. The annotation highlight regions where the HPF image has fewer noise as compared to the LSRTM image. The HPF iterative migration image was generated using 20 iterations, a weighting value of $\epsilon = 0.005$ and a model-styling threshold of $g_m = 25$.

Figure 10a shows one shot gather from the input data. After 1 iteration, Figure 10b and 10c show the data residual from LSRTM and HPF iterative migration, respectively. Figure 10e and 10f show the corresponding data residual at iteration 20.

Figure 11 shows the normalized power spectrum in the z-direction between the RTM, the LSRTM, and the HPF iterative migration images. Notice that, with iterative inversion, the LSRTM and HPF spectrum becomes more white, especially in the higher and lower frequency range. There are higher frequency noise in the LSRTM, which are better suppressed with HPF iterative migration. We can see that the HPF image has an overall flatter spectrum as compared to the LSRTM image. The relative amplitude of the HPF image in the range of $kz = 0.004 - 0.012m^{-1}$ is higher than that from LSRTM.

DISCUSSION

Overall, we oberve that the HPF iterative migration can suppress migration artifacts in the image. However, the spareness constraint tends to make some reflectors less continuous. One way to improve on the current result is to add a geophysical regularization term that enforces the continuity of amplitude along the reflectors. This require an estimation of the reflector's dip everywhere in the image and additional tuning parameters for the regularization term. A companion paper by Ma et al. (2014) has investigated the affects of including geophysical regularization in the iterative migration problem.

CONCLUSION

We compare inversion results of iterative migration using the classical least-squares norm and the hyperbolic penalty function. Reults from 2D field data examples show that using hyperbolic penalty function with model-styling can enforce sparsness in the image model. The sparse-reflectivity images obtained from hyperbolic penalty function iterative migration have less artifacts and better defined reflectors as compared to the LSRTM image. However, reflectors from the hyperbolic penalty function image are often less continuous and the parameterization of the hyperbolic penalty function remains cumbersome.

ACKNOWLEDGMENTS

Antoine Guitton acknowledges GeoImaging Solutions Inc. for financial support. The authors thank ExxonMobil for generously donatinng to SEP the field data used for the examples in this paper.

REFERENCES

Claerbout, J. F., 2014, Geophysical image estimation by examples.

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

Li, Y., B. Biondi, R. Clapp, and D. Nichols, 2014, Wave-equation migration velocity analysis for VTI models: Geophysics, **79**, WA59–WA68.

Ma, Y., M. Maharramov, and B. Biondi, 2014, Geophysical application of hyperbolic and hybrid L1/L2 optimation: SEP Report, **155**.

Valenciano, A., 2008, Imaging by Wave-Equation Inversion: PhD thesis, Stanford University.

Wong, M., B. Biondi, and S. Ronen, 2012, Imaging with multiples using linearized full-wave inversion: SEG Expanded Abstracts, 1–5.

Wong, M. and A. Guitton, 2014, Preliminary results of iterative 1D imaging with the hybrid penalty function: SEP Report, **152**.

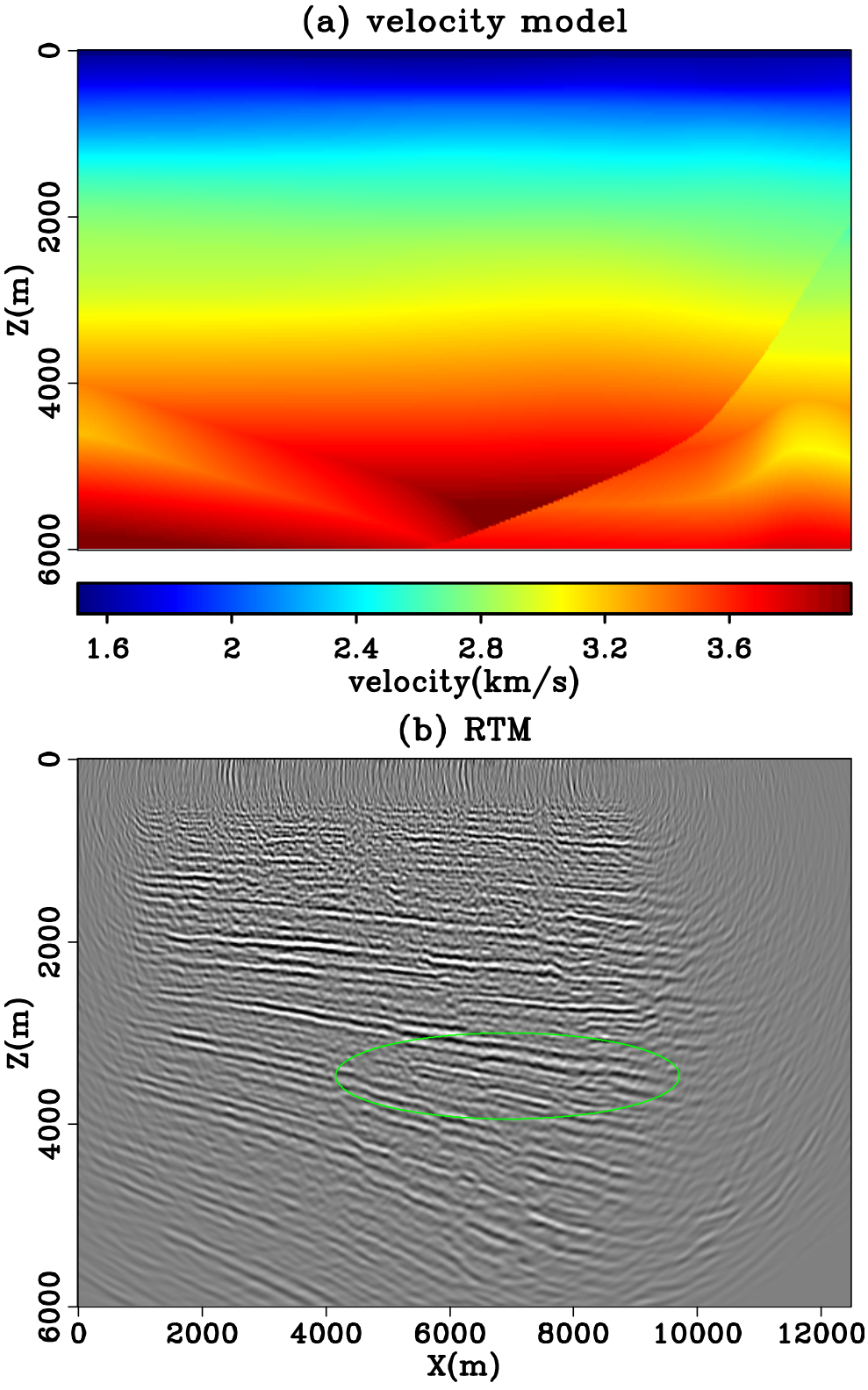


Figure 2: (a) The migration velocity model and (b) the RTM image. [**CR**] mandy1/. fig2-velrtm

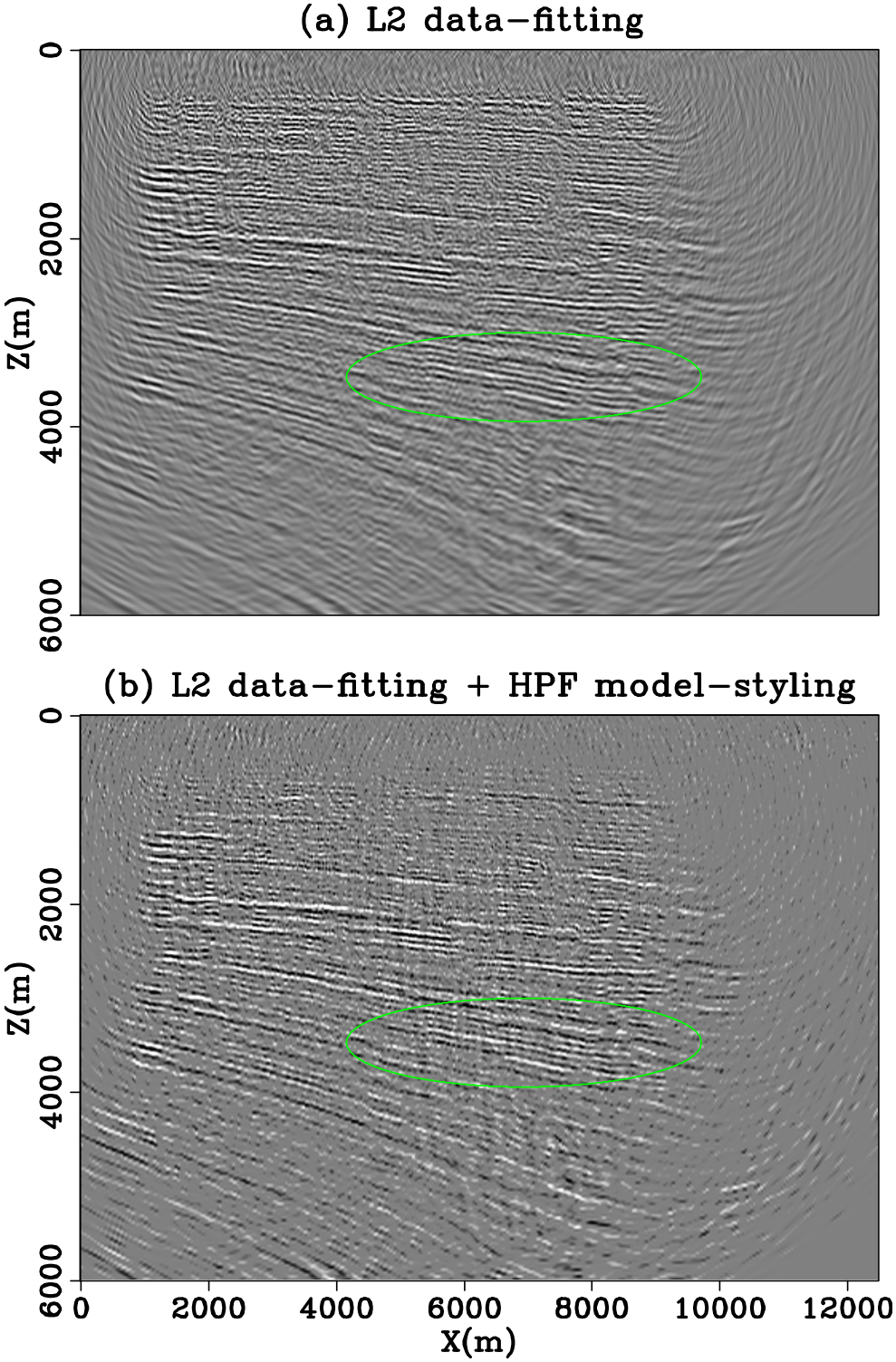


Figure 3: (a) LSRTM image using the ℓ^2 data-fitting objective function (equation 2) after 12 iterations of conjugate gradient algorithm. (b) The HPF iterative migration image using the ℓ^2 data-fitting with HPF model-styling objective function (3) after 20 iterations. [**CR**] mandy1/. fig3-inv

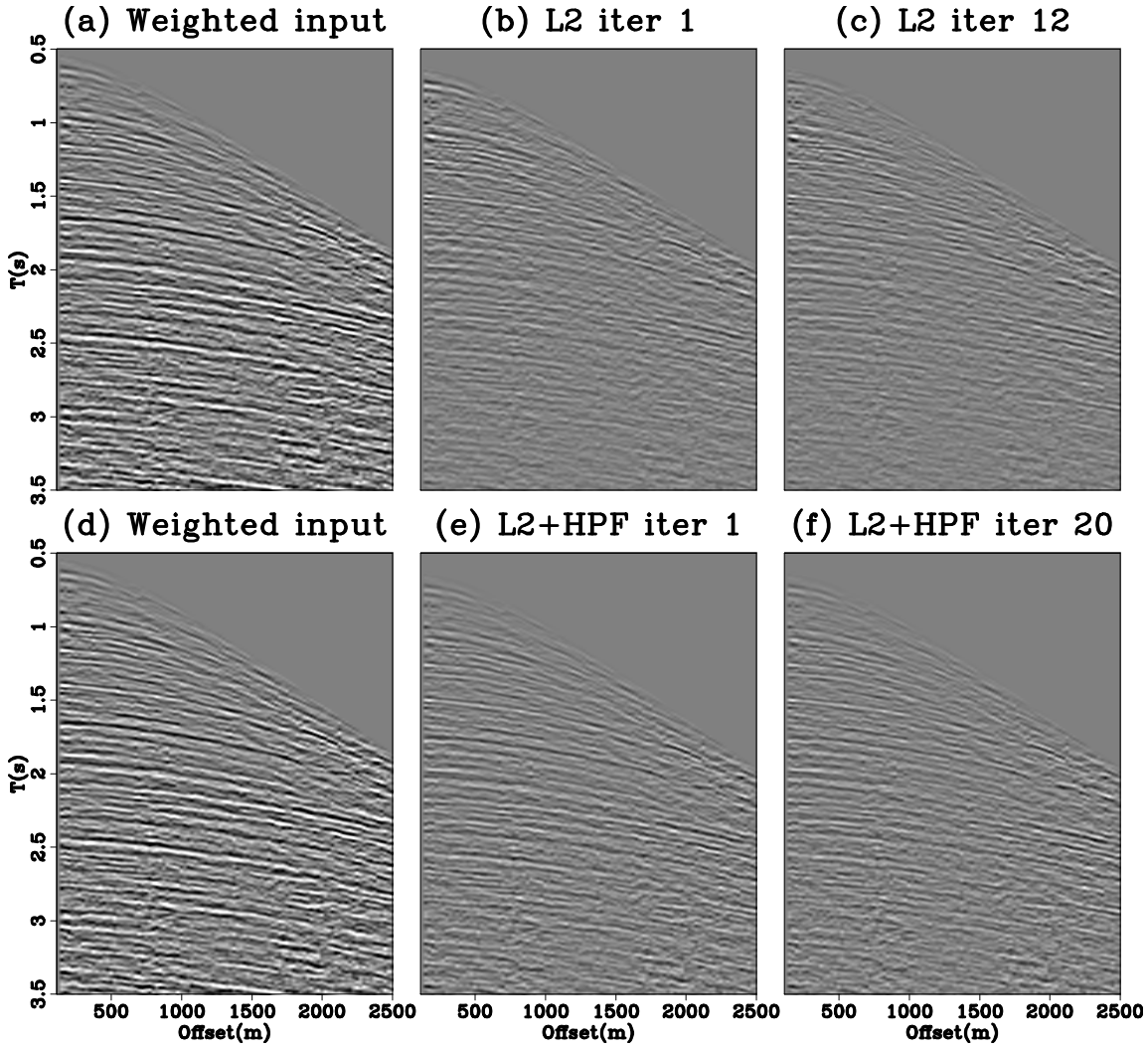


Figure 4: (a) and (d) shows one shot gather from the input data. The data residual from the same shot gather after (b) 1 iteration and (c) 12 iterations from LSRTM. The data residual from the same shot gather after (e) 1 iteration and (f) 20 iterations from the HPF iterative migration. All figures are clipped at the same level. [**CR**] mandy1/. fig4-resid

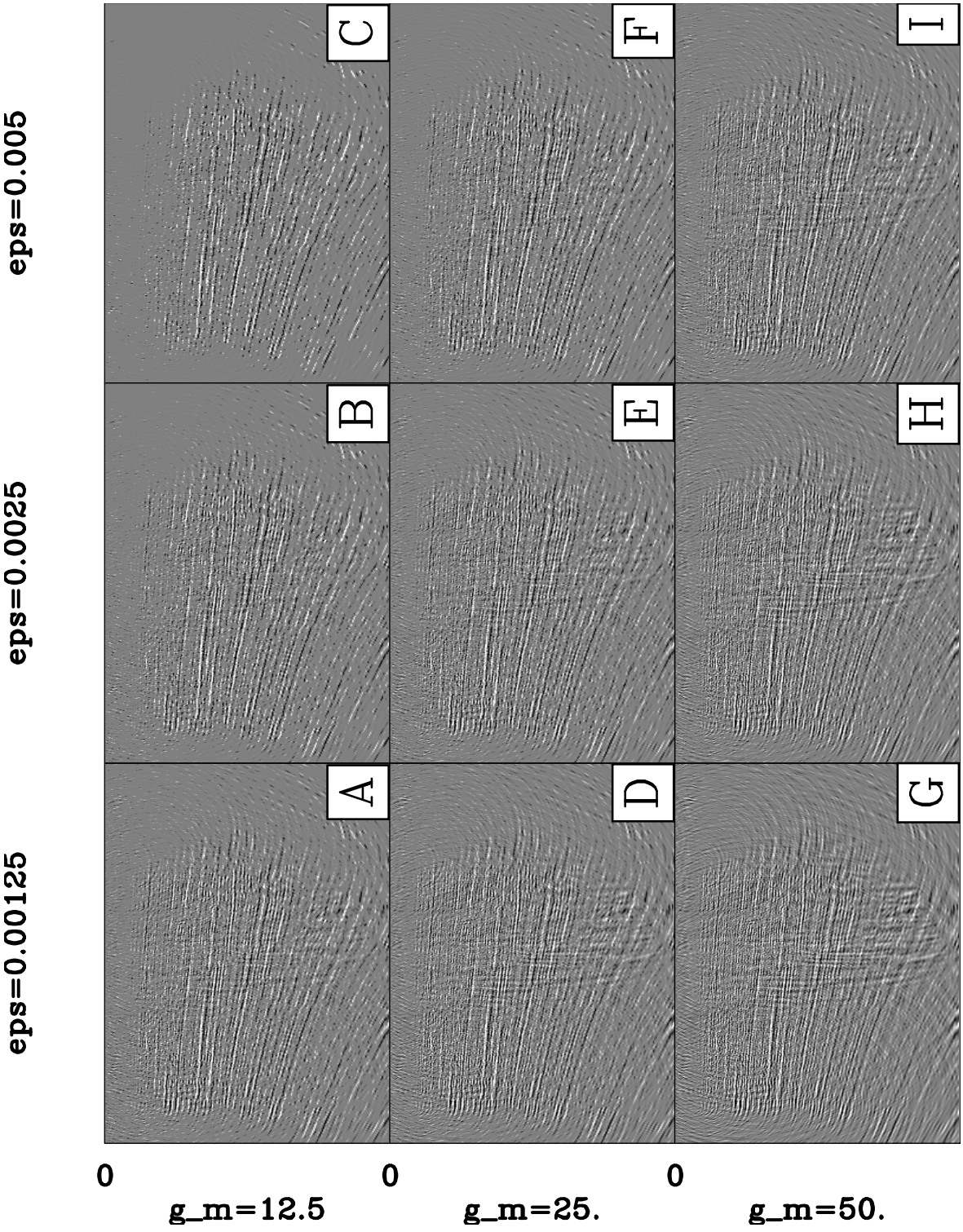


Figure 5: Inverted images from the HPF iterative migration using different ϵ and model-styling threshold values (g_m). Images from a common column use the same ϵ while images from a common row uses the same g_m. Panel A was selected as the HPF iterative migration result in Figure 3. [**CR**] mandy1/. fig5-3by3

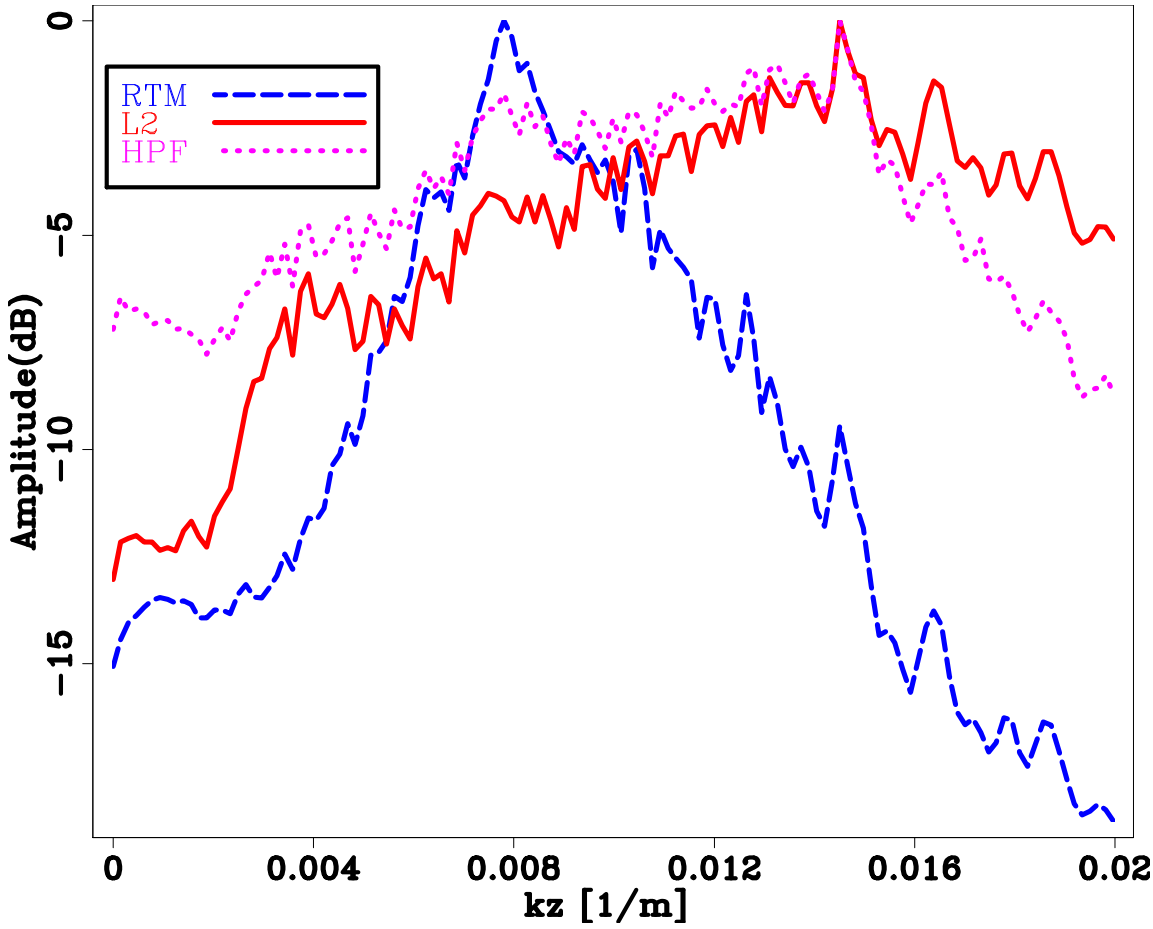


Figure 6: The normalized power spectrum in the z-direction for the RTM image, LSRTM image, and HPF iterative migration image from the first dataset. Iteration inversion raise the higher and lower end of the spectrum as compared to migration. [**CR**] mandy1/. fig6-power-sedi

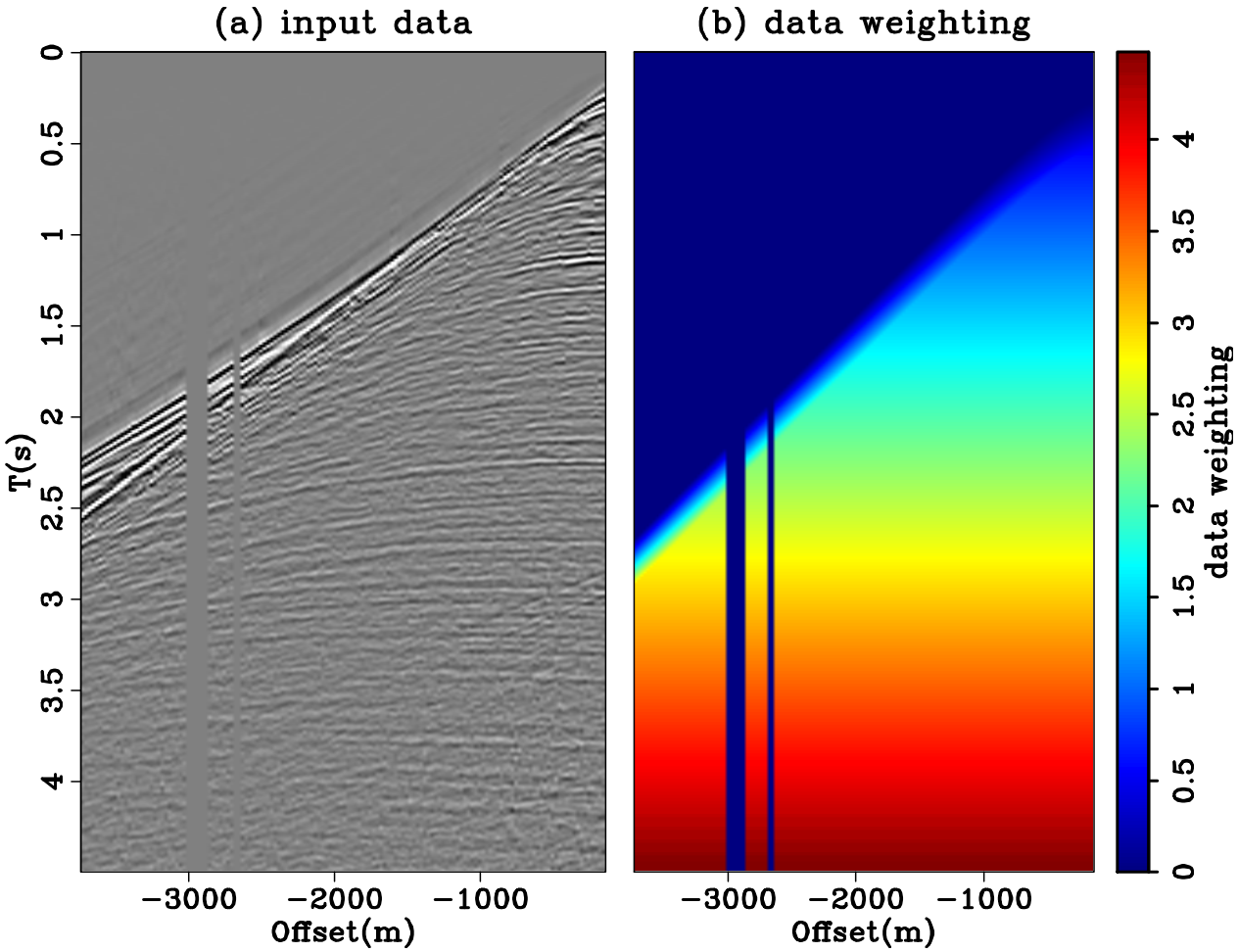


Figure 7: Input data for iterative migration in the salt region. (a) is a shot gather. The amplitude of the earlier reflection is much stronger than the deeper reflection. Therefore, (b) shows the data-weighting function used in the inversion to boost up the energy from the deeper part of the subsurface. [**ER**] mandy1/. fig7-inputdata-salt

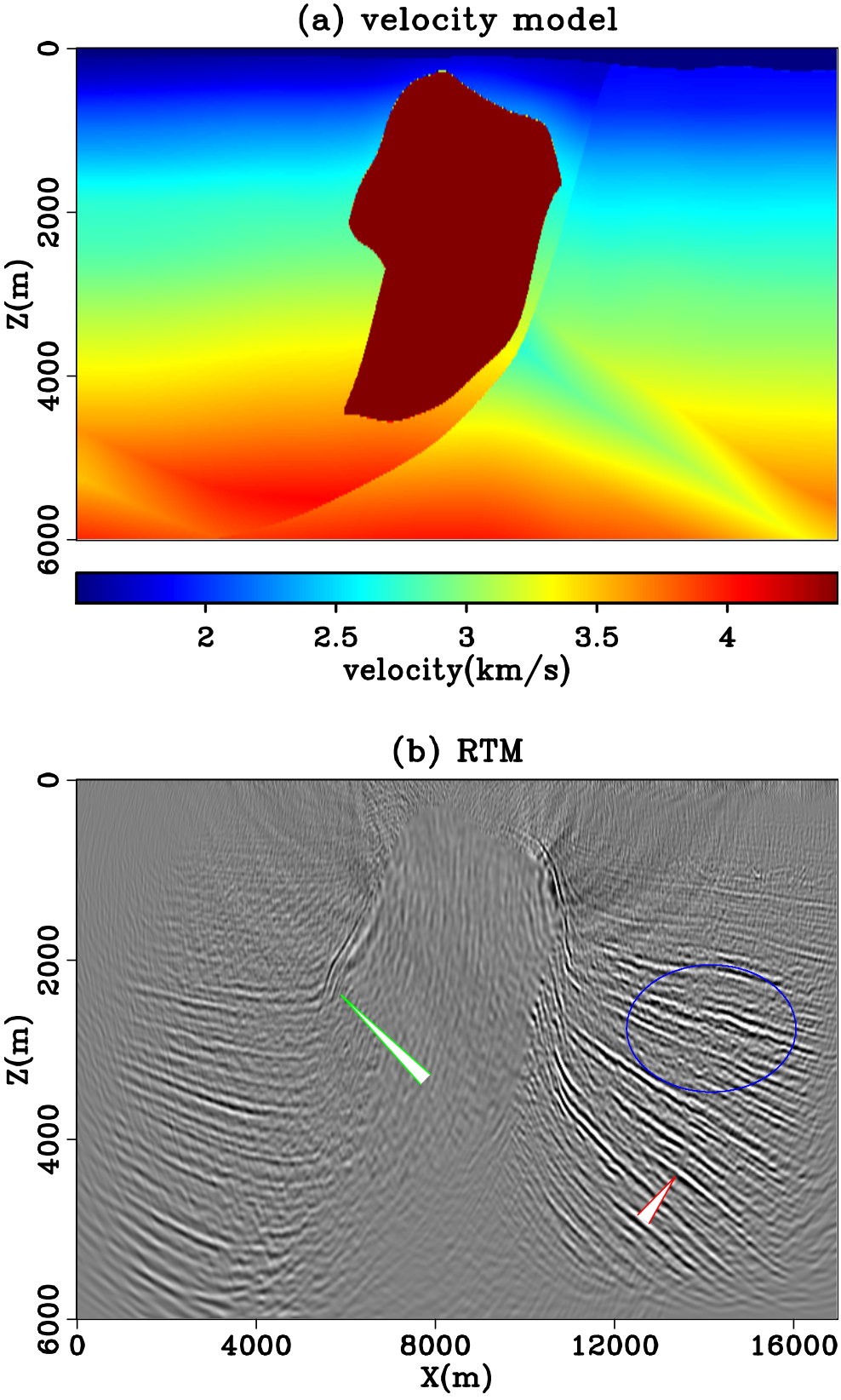


Figure 8: (a) The migration velocity model and (b) the RTM image. [**CR**] mandy1/. fig8-velrtm-salt

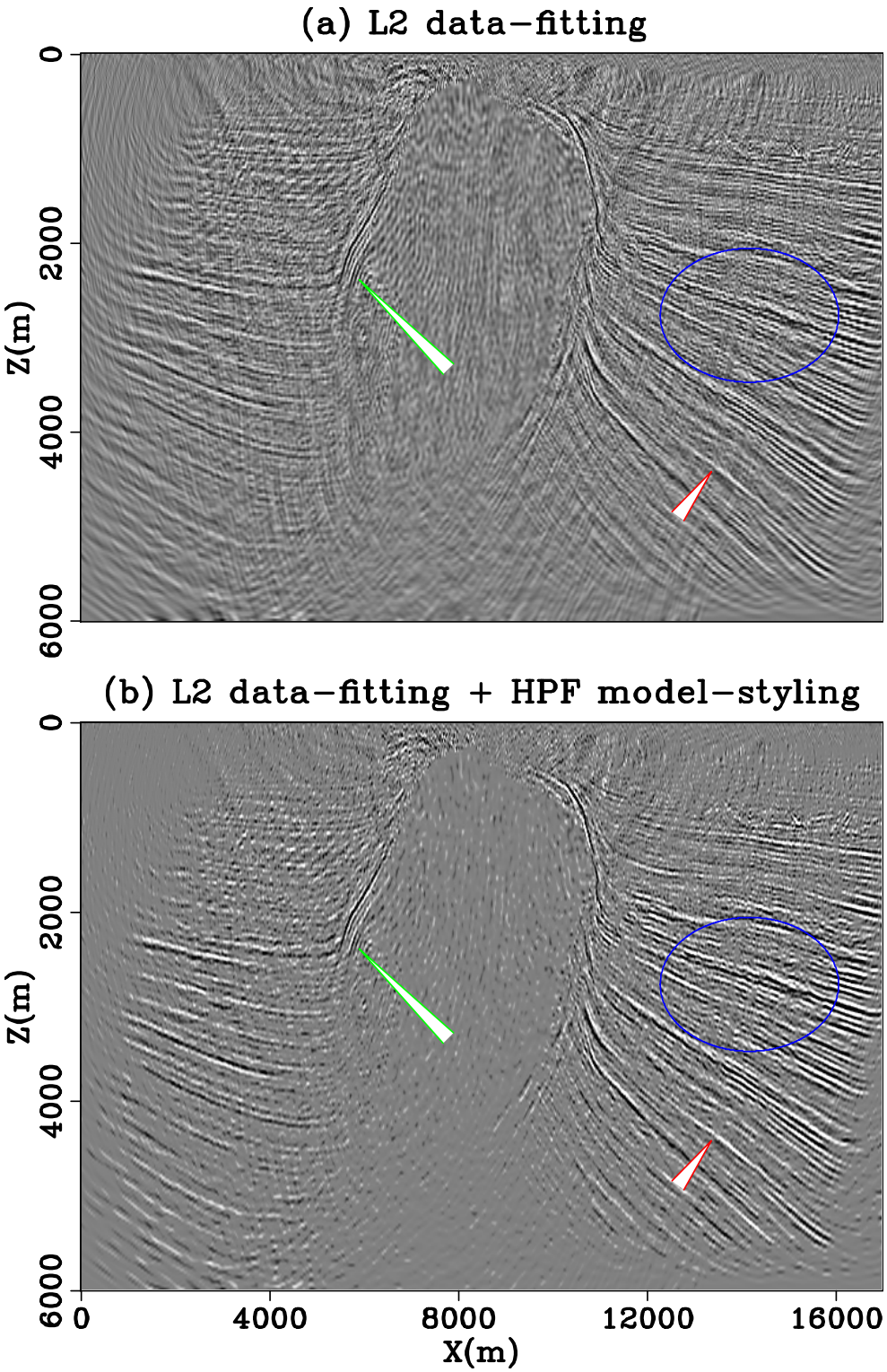


Figure 9: (a) The LSRTM image using the ℓ^2 data-fitting objective function (equation 2) and (b) the HPF iterative migration image using the ℓ^2 data-fitting with HPF model-styling objective function (3). Both inversions uses 20 iterations of the conjugate gradient algorithm. [**CR**] mandy1/. fig9-inv-salt

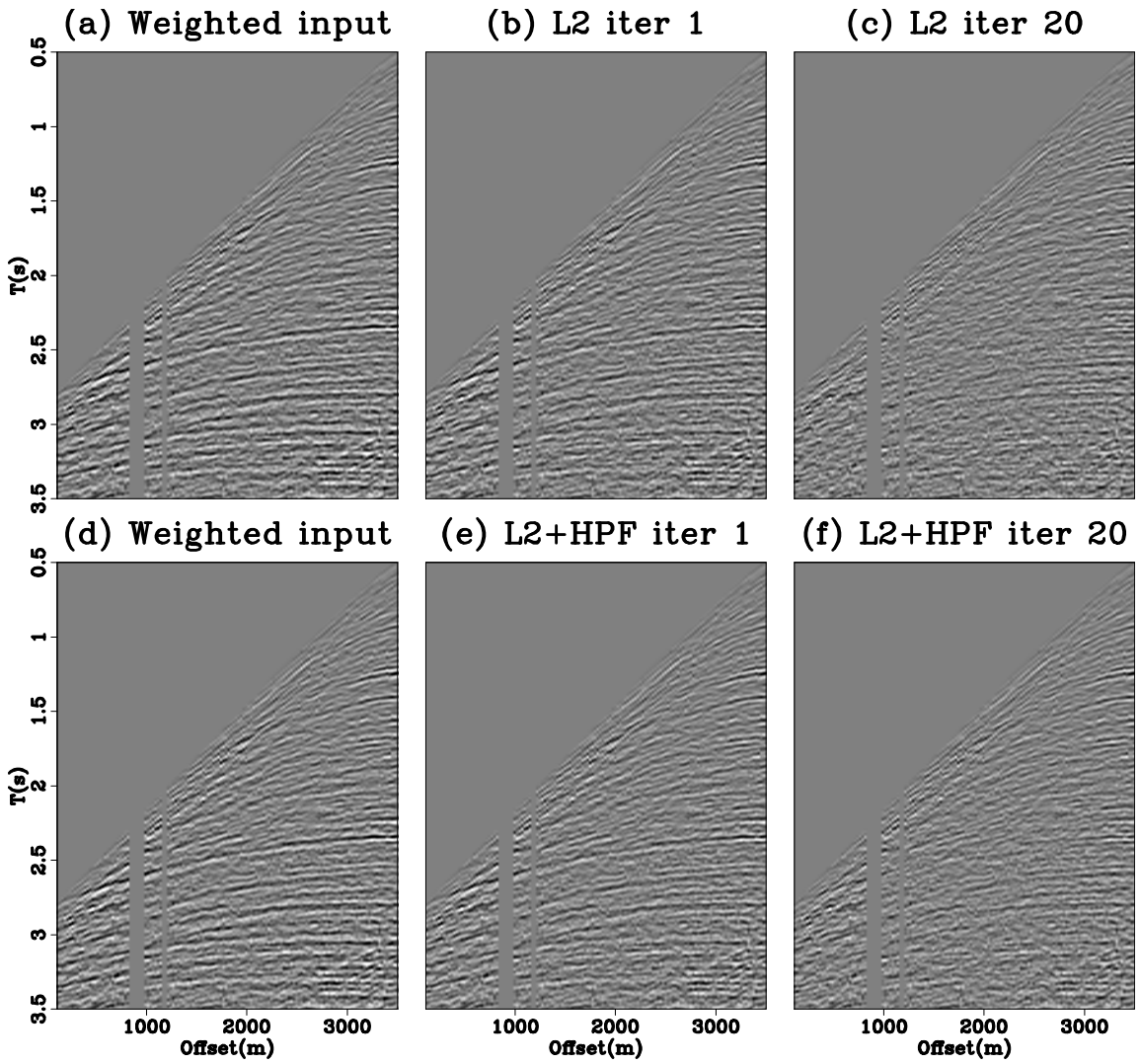


Figure 10: (a) and (d) shows one shot gather from the input data. The data residual from the same shot gather after (b) 1 iteration and (c) 20 iterations from LSRTM.The data residual from the same shot gather after (e) 1 iteration and (f) 20 iterations from the HPF iterative migration. All figures are clipped at the same level. [**CR**] mandy1/. fig10-resid-salt

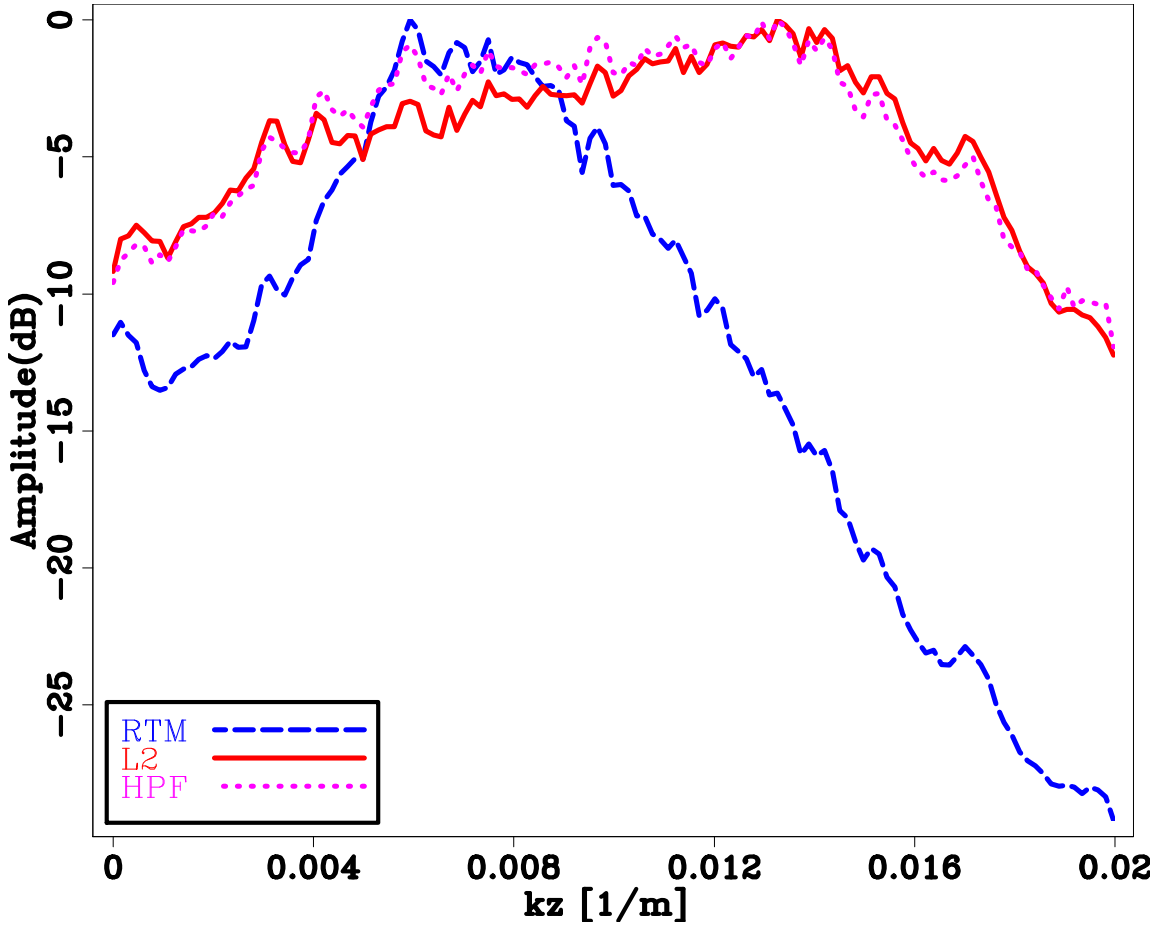


Figure 11: The normalized power spectrum in the z-direction for the RTM image, LSRTM image, and HPF iterative migration image from the second dataset. [**CR**] mandy1/. fig11-power-salt

Application of the up-down separation using PZ calibration filter based on critically refracted waves

Ettore Biondi and Stewart A. Levin

ABSTRACT

On marine multicomponent data we may apply PZ summation to separate the up-going and down-going wavefields. We show an example of this procedure for an ocean bottom nodal (OBN) dataset, where we perform an acoustic decomposition that gives the up- and down-going pressure fields in the water. To design the necessary calibration filters, we target long-offset refracted waves that contain purely up-going energy. After the acoustic separation, we adaptively subtract the down-going wavefield from the original pressure component. The quality of the separation achieved demonstrates that the calibration filters obtained from the long-offset waveforms were effective in decomposing almost all the events present in each gather.

INTRODUCTION

When dealing with multicomponent ocean bottom acquisition (OBN), where both the particle velocity vector and pressure are measured, a fundamental preprocessing step is wavefield up-down separation. In the context of wavefield separation of ocean bottom data, there are two possibilities (Schalkwijk et al., 1999). If we perform an acoustic decomposition, we obtain the down-going pressure just above the ocean bottom, and the up-going pressure just beneath the ocean bottom surface. Alternatively, with an elastic decomposition, we separate the up-going and down-going P- and S-waves just beneath the ocean bottom.

After such wavefield decomposition of the OBN data, the result can be used in further processing. For instance, source signatures can be estimated from the down-going wavefield after source static corrections are applied (Wong and Ronen, 2009). It is also possible to apply up-over-down deconvolution to attenuate all free-surface multiple (Grion, 2010). Furthermore, having a decomposed pressure wavefield leads to a better mirror imaging or a joint least-square inversion of up- and down-going signals (Grion et al., 2007; Wong et al., 2010). These examples are only few applications where we need the decomposition of the wavefield, underscoring the importance of doing this processing step when dealing with ocean bottom acquisition.

There are many different published methods for performing PZ summation (Sonneland et al., 1986; Barr and Sanders, 1989; Amundsen, 1993). In our application, we proceed with an acoustic decomposition of the pressure field performed in the physical (t-x) domain of an OBN acquisition. Usually, there are differences in coupling and instrument response of the measuring devices. Therefore, the data must be properly calibrated before the pressure (P) and the vertical velocity (Z) components can be combined. We follow the PZ summation algorithm developed by Melbø et al. (2002) where the necessary calibration filter is computed by minimizing the energy of the resulting down-going wavefield within a manually picked

window tracking critically refracted waves. We start showing the result of the elimination of the instrument response of both the pressure and vertical velocity components. Then, we explain how the PZ summation algorithm works, and show the results. Finally, we present the result of the adaptive subtraction procedure developed by Alvarez and Guitton (2007) subtracting the down-going pressure obtained from the PZ summation method to the original P component. We applied the acoustic decomposition to all the receivers of the line. Here, we report the result of a representative common-receiver gather belonging to the inner portion of the recording line. Our result demonstrates that the calibration filter based on the long-offset refracted arrivals can be used to separate the energy for all the offset range.

RECEIVER IMPULSE RESPONSE DESIGNATURE

The OBN dataset was recorded in the North Sea and appears in our data library under the name Moere Vest. We worked with a 2D survey composed of 179 multicomponent receivers with a minimum receiver interval of approximately 250 meters (m) and a maximum of 10 kilometers (km). The source interval was approximately 100 m. The inner portion of the line has the fine receiver interval; only at the two ends does the increase to approximately 10 km (see Alves (2014) for all the dataset details).

Figure 1 shows the P and Z components of the representative common-receiver gather, bandpassed to the frequencies contained in the sources. In this figure, you can clearly see the direct arrival and its free-surface multiples, indicated by the green arrows, with periodicity of approximately 2 seconds (s). On this dataset, a debubbling procedure has not been applied and source reverberations are evident. Thanks to the available long offsets, it is possible to observe refracted waves starting at approximately 8 km offset. It is interesting to note the sea-surface multiple of the first refracted arrival centered around 7.5 s on the right side of both components. In addition to these events, a clear primary reflection appears at 2.5 s (pointed by the red arrow). All these events help us understand the result of the wavefield decomposition because they can be easily divided between up- or down-going arrivals. Both components contain the same events except for some water reverberations and shear-wave induced events present in the recorded vertical velocity. However, the different instrument responses cause the equivalent arrivals in the two components to have amplitude and phase differences (compare the direct arrival and its first multiple or the primary reflection at 2.5 s). To compensate for these differences, it is necessary to correct for the different impulse responses of the measuring devices.

Fortunately, we were provided with impulse responses of the deployed hydrophone and geophone. Figure 2a displays the time impulse responses for the hydrophone and the geophone, on the right and left respectively. One curious feature of these responses is their noncausality, which is caused by the correction for the linear phase shift introduced by the the antialiasing filter during the recording of the data. Looking at the amplitude spectra responses (Figure 2b), we see that the effect of the receiver impulse response designature boosts the low frequencies for both components. Regarding the phase instrument spectra, shown in Figure 2c, it is possible to see that the phase correction for the geophone affects wavelet shapes more then the correction for the hydrophone. Figure 3 shows the comparison of the amplitude spectra before and after the receiver impulse response correction for both components. In the case of the hydrophone, the frequencies up to 4 Hertz (Hz) have been

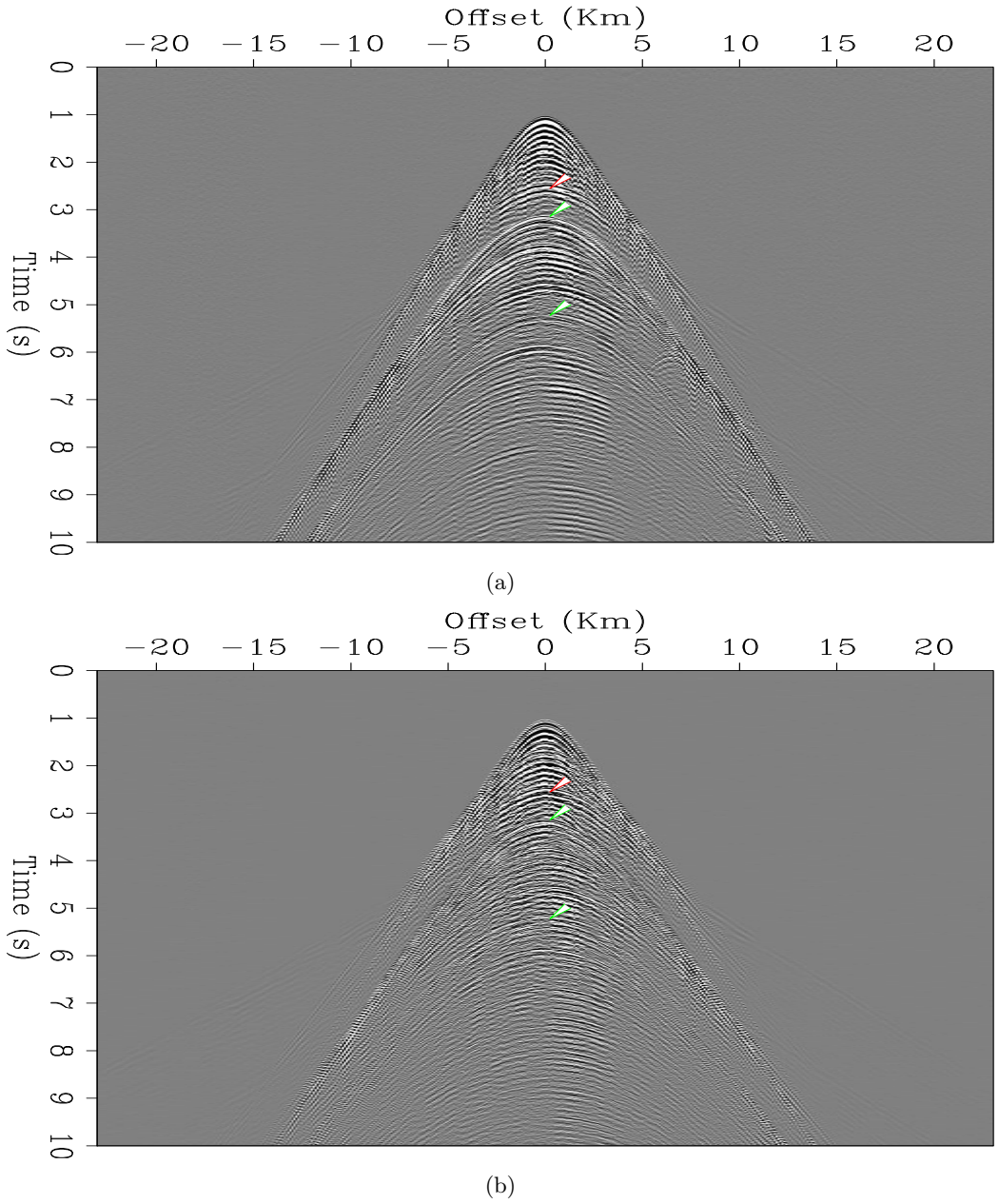


Figure 1: Portion of a common-receiver gather of the Moere Vest OBN dataset. (a) Pressure component of the data. (b) Recorded vertical velocity. In both components there are visible free-surface multiples of the direct arrival with a periodicity of approximately 2 s (indicated by the green arrows). Focusing attention on the near-offset events, it is possible to see one primary reflections centered around 2.5 s (indicated by the red arrow). As expected, the vertical component contains the same events as the recorded pressure, except for a very low energy water reverberations and shear-wave induced noise. However, looking at the gathers carefully, it is possible to see amplitude and phase differences between equivalent arrivals in the two components. This effect is caused by the different instrument response and coupling of the measuring devices. [**ER**] ettore1/. orighyAnt,origzAnt

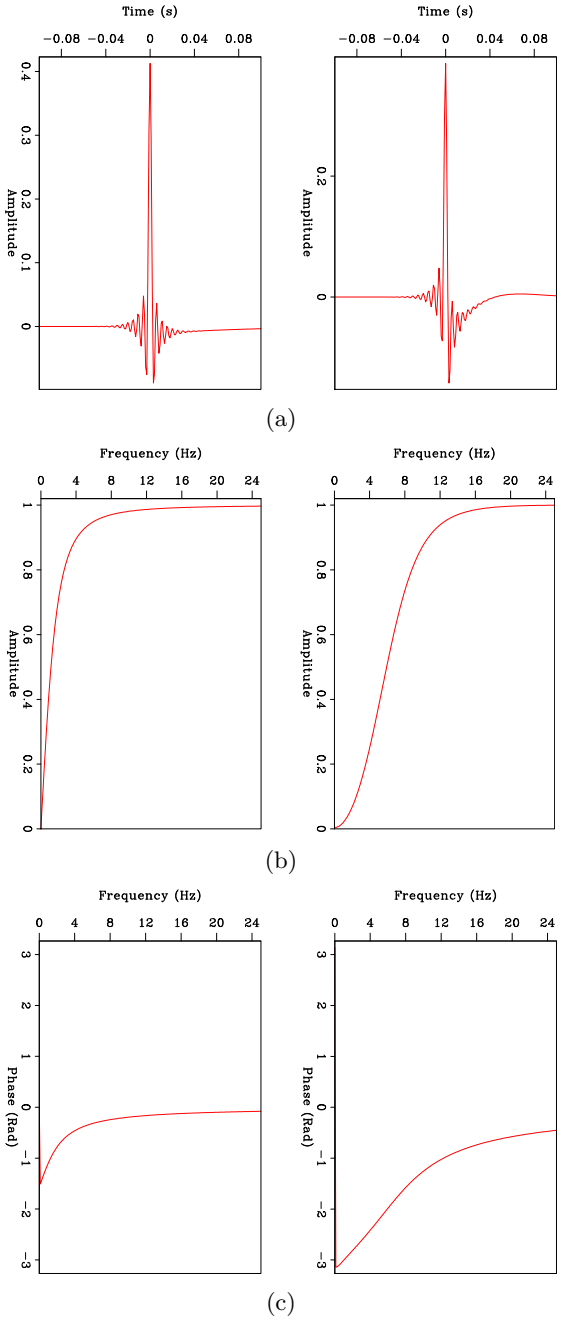
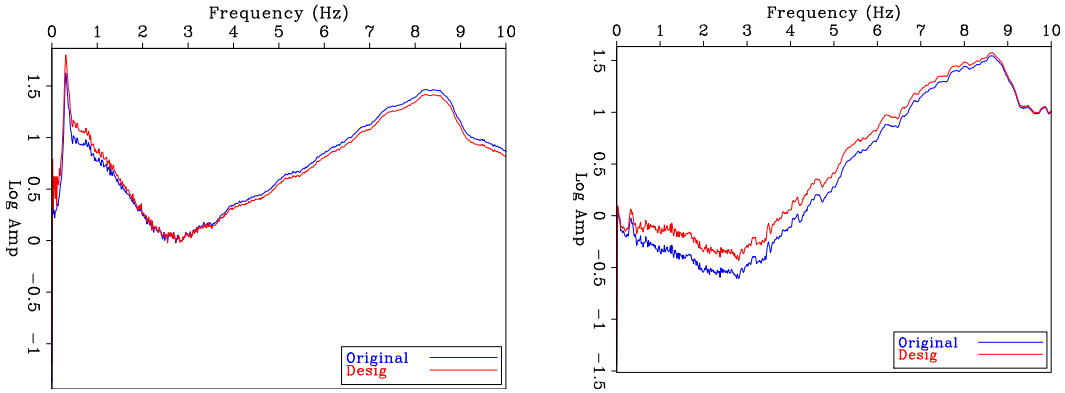


Figure 2: Comparison of the instrument responses of the hydrophone (right panels) and of the geophone (left panels). (a) Instrument responses in the time domain. (b) Amplitude spectra of the responses of the measuring devices. (c) Phase spectra of the instrument responses. It is interesting to point out the noncausality of both instrument responses. This nonphysical feature is present because the data has been previously corrected for the linear phase shift of the antialiasing filter (approximately 60 ms), and it has been taken into account in the computed instrument responses. [**ER**] ettore1/. time,amp,phase



(a) (b)

Figure 3: Comparison between the amplitude spectra of the pressure (a) and vertical (b) components before and after the instrument impulse response designature. The receiver designature boosts the low frequencies in both the components of the data. [**ER**] ettore1/. spectracomparehy,spectracomparez

noticeably boosted; while for the geophone, the correction has enhanced frequencies up to approximately 9 Hz.

Comparing equivalent events present in both the components of the representative common-receiver gather after the receiver response designature, affirms that indeed, this preprocessing procedure eliminated most of the amplitude and phase differences between the recorded P and Z (Figure 4). Now, the P and Z components behave according to what theory predicts: the direct arrival and free-surface multiples in the pressure have opposite polarity respect the same events in the Z component. Instead, the refracted first breaks and primary reflections have the same polarity on the two components (e.g., the reflection centered at 2.5 s).

UP-DOWN SEPARATION THROUGH PZ SUMMATION

Having compensated for the different instrumental impulse responses, we now have to correct the data for the coupling effect that still affects our data. This correction is conducted by means of a calibration filter $a(f)$ applied to one of the two recorded components, which in our case, is the vertical velocity. Using this filter for performing the acoustic decomposition, the up- and down-going pressure wavefields can be expressed as follows:

$$P_{up}(f,k) = \frac{1}{2}P(f,k) + a(f)\frac{\rho}{2q(f,k)}Z(f,k), \tag{1}$$

and

$$P_{down}(f,k) = \frac{1}{2}P(f,k) - a(f)\frac{\rho}{2q(f,k)}Z(f,k), \tag{2}$$

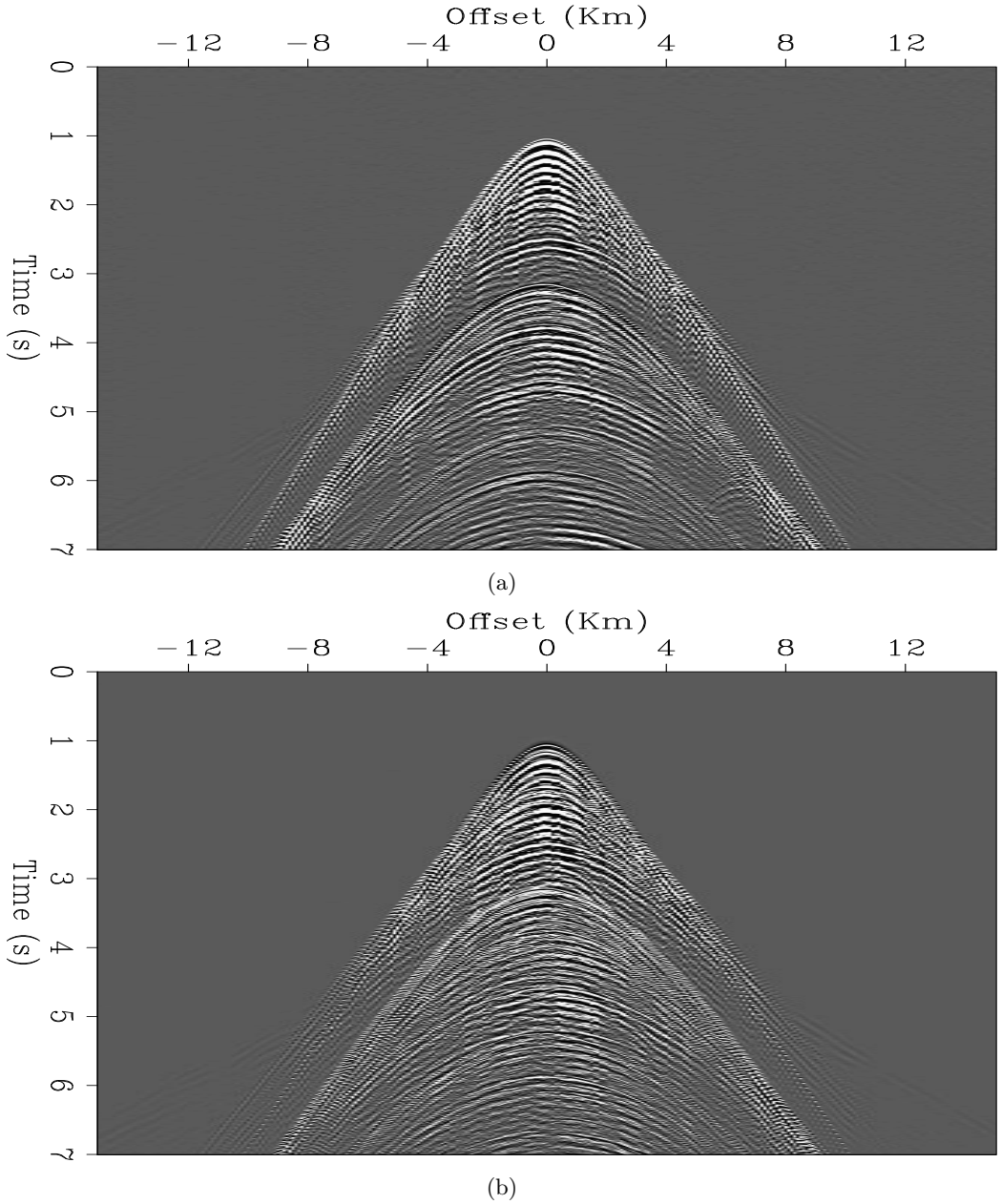


Figure 4: Close-ups the common-receiver of Figure 1 after the the receiver response designature. The correction for the different responses of the instruments makes the up-going events (e.g., refracted and diving waves) in the pressure (a) and the vertical (b) components more similar, and down-going events (e.g., the direct arrival and free-surface multiples) to have the opposite polarity. [**ER**] ettore1/. desighy,desigz

respectively, where P is the pressure, f is the frequency, k is the horizontal wave-number, Z is the vertical velocity component, ρ is the density of the water, and q is the vertical slowness in the water layer. This last factor is related to the ray parameter p with the following relation:

$$q(f,k) = \sqrt{c^{-2} - p^2(f,k)}, \tag{3}$$

where c is the velocity of the water at the receiver position. These equations hold for common-shot gathers, but it is possible to use them for decomposing wavefields in common-receiver gathers (Schalkwijk et al., 1999). The calibration filter $a(f)$ is found by minimizing the energy of the down-going pressure (Eq. [2]), in a least-squares sense, in a window containing only up-going events.

In our application, we manually picked a window tracking refracted waves before the first refracted multiple events. In particular, we picked a window of approximately 1 s of the long-offset refracted waves between 11 km and 15 km. To find the calibration filter, we computed a Wiener shaping filter that shapes the Z component wavelets to the wavelets of the pressure wavefield within the picked window. As a result, such a filter also incorporates the density and slowness factors of Equations (1) and (2). The density factor does not vary for all the receiver gather. Instead, the vertical slowness varies both in frequency and wave-number (or time and offset in the physical domain). Inside the window, we can assume that the slowness factor to be constant, and we can easily compute it using the slope of the first break refracted arrivals. To simplify the computation of the varying vertical slowness in the rest of the gather, we calculated the local slowness of the events inside the gather using a constant water velocity model with the vertex of the first hyperbola centered at the apex time of the direct arrival. Above the direct arrival, we use the slope of the refracted waves for all the other events. In addition, the vertical slowness is considered to be local, and such assumption allows us to use it as a multiplicative factor in the physical domain. Without this assumption, it would be a bidimensional convolution in the offset-time domain.

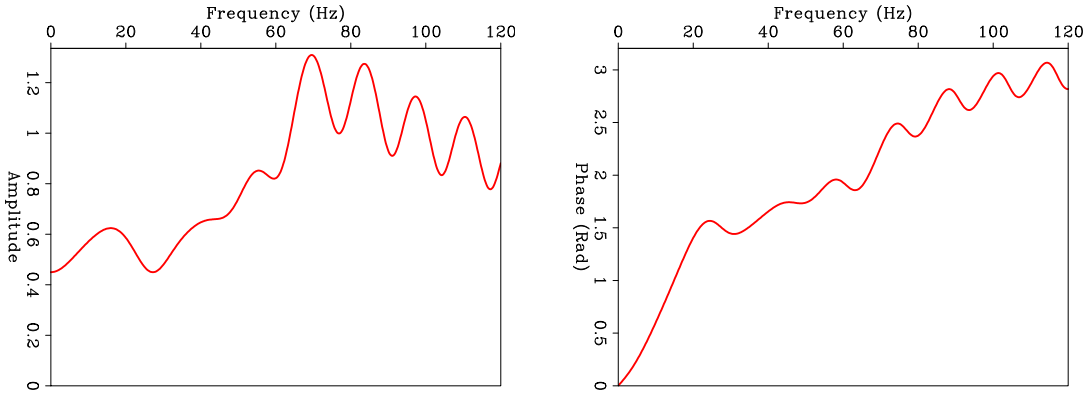


Figure 5: Frequency response of the calibration filter with 41 coefficients obtained from the long-offset refracted waves. (a) Amplitude spectra of the filter. (b) Phase spectra of the filter. [**ER**] ettore1/. filteramp,filterphase

Figure 5 shows a plot of the amplitude and phase frequency response of the calibration filter with 41 coefficients. The plot clearly shows that the filter varies with frequency. Therefore, the coupling differences between the two components are not related to a simple scalar amplitude factor. In fact, the result obtained only with a constant calibration factor in Melbø et al. (2002) demonstrated the limitations of such white spectra filter.

In Figure 6, we display the result of the PZ algorithm after calibrating the Z component of the reported receiver with the filter of Figure 5. In the down-going wavefield, most of the up-going energy has been attenuated; indeed, the primary reflection at 2.5 s has almost completely vanished. Furthermore, the primary refracted events are separated and present in the up-going pressure wavefield; whereas, the free-surface multiples of the critically refracted waves are present only in the down-going component. In the up-going pressure, it is now possible to identify two distinct primary reflections, one at 1.9 s and the other at 2.5 s. The leftover down-going energy still present in the up-going wavefield is cause by the fact that we performed an acoustic separation where we are obtaining the up-going pressure just beneath the sea bottom (Grion, 2010).

To improve the result for the up-going wavefield, and, therefore decrease the leakage of down-going energy into it, we adaptively subtract the down-going pressure of Figure 6b from the original P component of Figure 1a. The adaptive subtraction was conducted with the algorithm developed by Alvarez and Guitton (2007) using a five-coefficient temporal matching filter applied to all the data in the gather. Comparing the up-going pressure obtained with the PZ algorithm with the same wavefield conducted using the adaptive subtraction (Figure 7), after performing the hyperbolic moveout (HMO) correction with water velocity, it is clearly visible that the adaptive subtraction routine further attenuated the direct arrival and its bubble pulses.

It is interesting to see the down-going pressure wavefield after applying the HMO correction with water velocity (Figure 8). The direct arrival and its bubble become flat and easily detectable. Although some energy of the primary events leaks into this wavefield, because of the simple slowness model and the local approximation used, it is possible to distinguish up to 12 pulses of the source. This wavefield is indeed a point of departure for the source estimation and further processing, such as the debubbling of the data.

To evaluate the quality of the wavefield separation for the other receivers, in Figure 9 we show the costant-offset section at approximately 0 km offset. In the up-going pressure field section, obtained with the adaptive subtraction algorithm, the direct arrival and its reverberations are attenuated for all the shown receivers. The two primary reflections previously highlighted are evident and it is possible to follow their trend along the acquisition line. In addition, in the down-going wavefield section the free-surface multiples of the two reflections are visible, albeit they have a lower amplitude.

CONCLUSIONS

We performed an acoustic decomposition of the wavefield of OBN data using a PZ summation algorithm based on calibration filters computed from the critically refracted waves. To compensate for the different impulse responses and coupling of the recording devices, we first proceeded with an instrument response designature for both the components. After that, we picked a window containing only up-going refracted waves and design a shaping

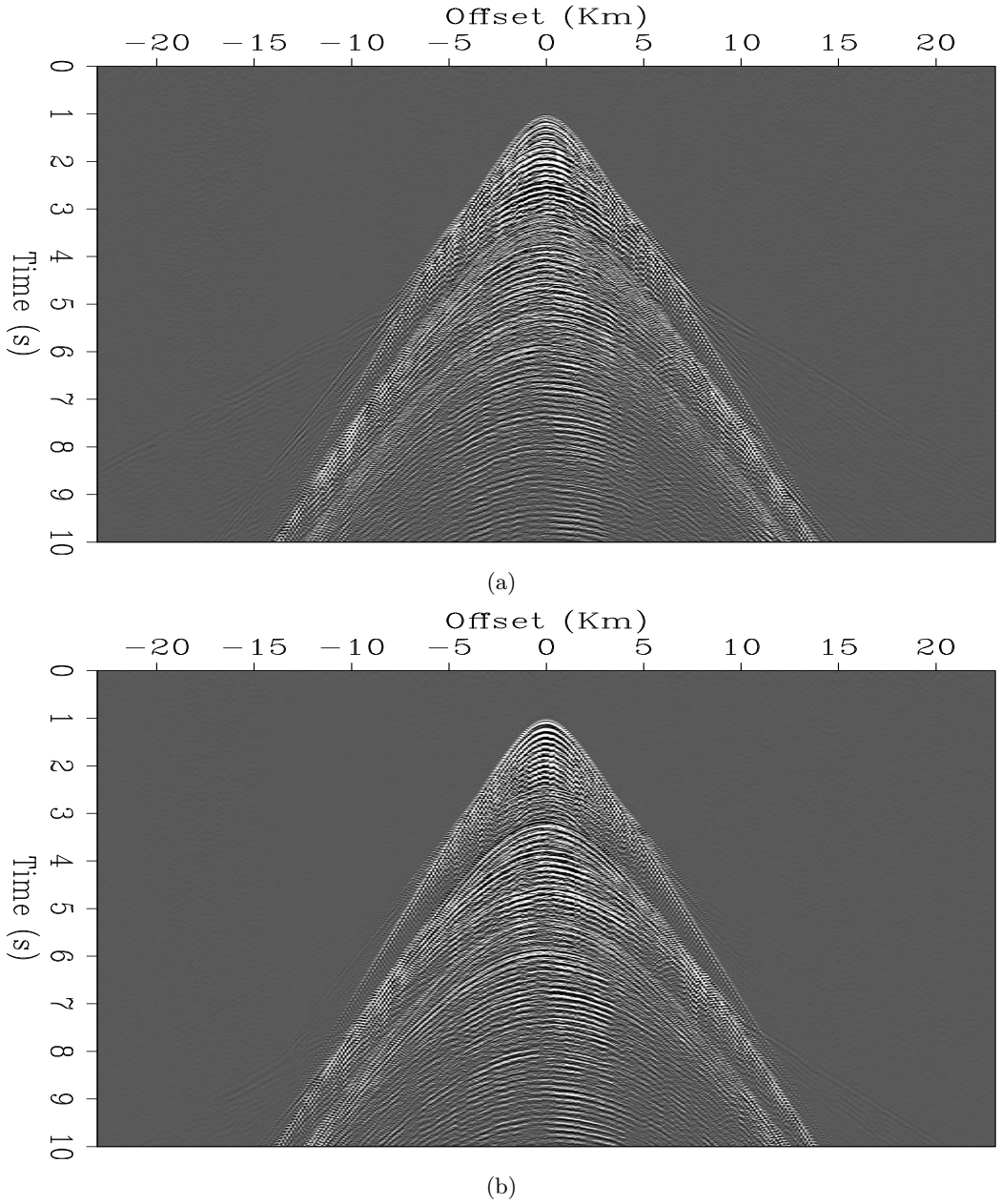


Figure 6: Output of the PZ summation having applied the calibration filter of Figure 5 to the vertical component. (a) Up-going pressure wavefield just beneath the ocean bottom. (b) Down-going just above the ocean bottom. Most of the energy is well separated, especially for the down-going wavefield where most of up-going refracted arrivals are attenuated. Note that the multiple of these events are present only in the down-going pressure. Because we are performing an acoustic decomposition, some of the energy of the down-going waves is still present in the up-going wavefield. Despite that, the two primary reflections centered about 1.9 s and 2.5 s are only present in the up-going pressure. [**ER**] ettore1/. up,down

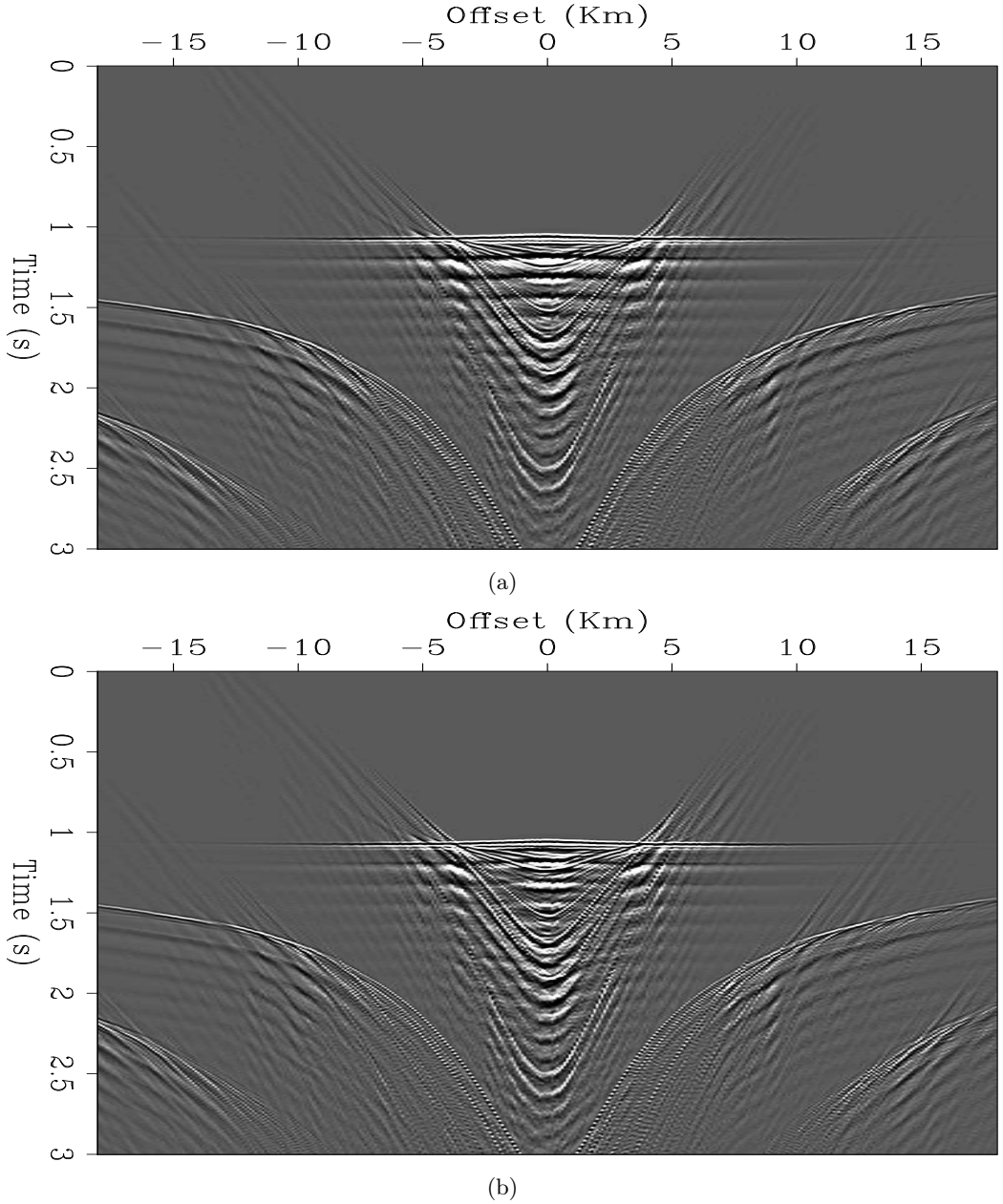


Figure 7: Result of the adaptive subtraction of the down-going wavefield of Figure 6b from the pressure component of Figure 4a after the hyperbolic moveout (HMO) correction with a water velocity of 1,475 m/s. (a) Close-up of Figure 6a after the HMO correction. (b) Resulting up-going pressure using the adaptive subtraction developed by Alvarez and Guitton (2007). In the up-going wavefield obtained with the adaptive subtraction routine, the direct arrival and its bubble is clearly attenuated. [**ER**] ettore1/. hmo,hmo1

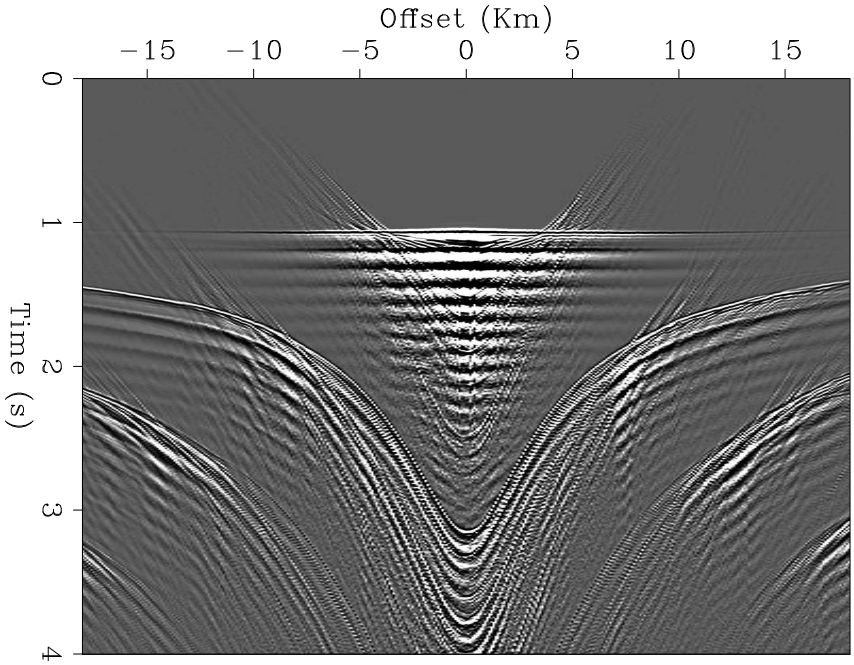


Figure 8: Close-up of the down-going pressure of Figure 6b after the application of the HMO correction. Looking at the direct arrival, it is possible to see 12 pulses of the source on the near offsets. The leakeage of up-going energy (e.g. refractions at about 4 km offset) it is caused by the simple slowness model and the local approximation considered in the PZ summation algorithm. [**ER**] ettore1/. downhmo

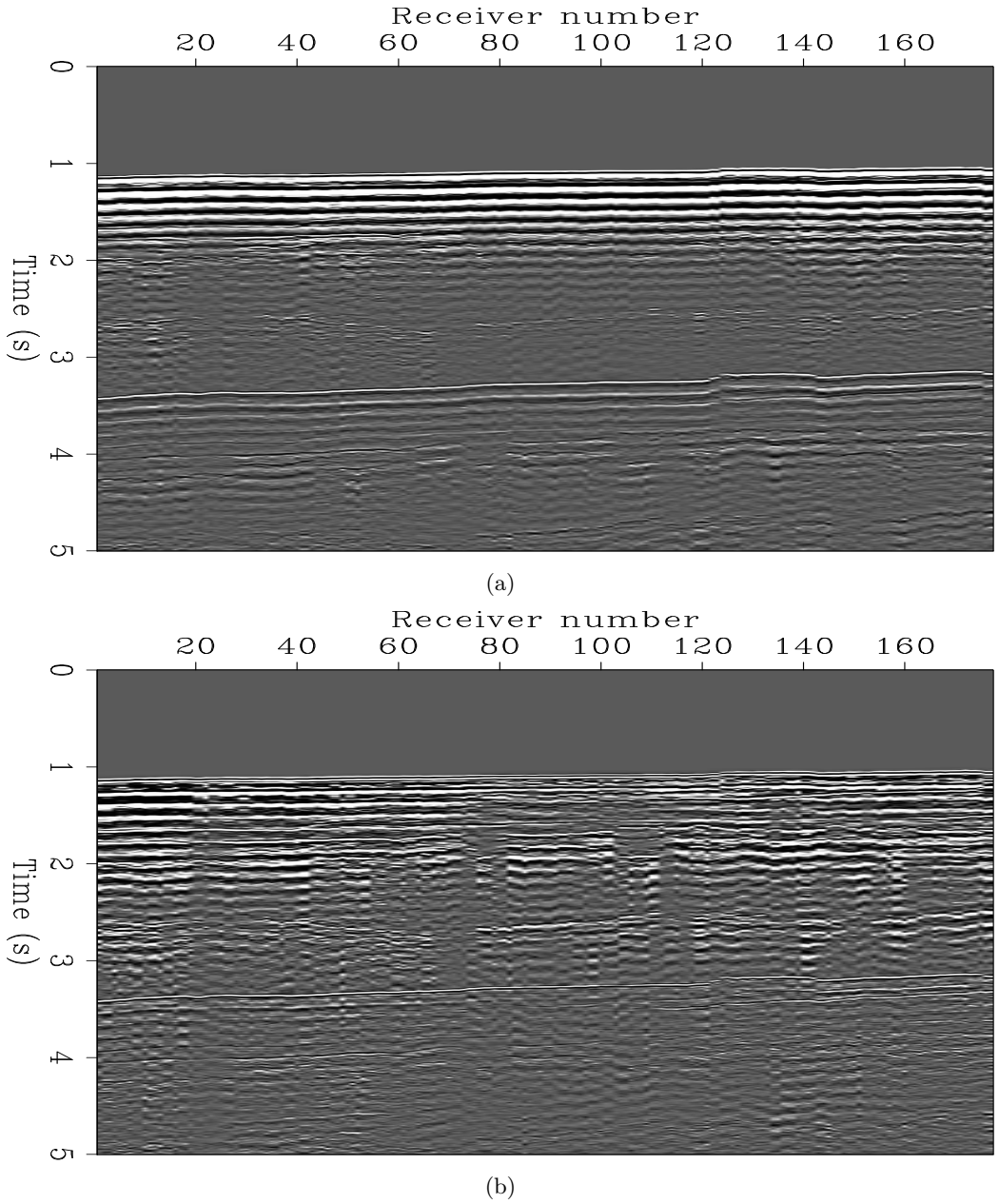


Figure 9: Comparison between constant-offset sections for the up- and down-going wavefields at approximately zero offset. (a) Constant-offset section of the down-going pressure field. (b) Constant-offset section of the up-going wavefield obtained with the adaptive subtraction. In the top panel, the two events around 1.9 and 2.5 are primary reflections. It is possible to see their trend along the acquisition line. Instead, in the bottom panel the direct arrival and its first free-surface multiple are clearly visible. In addition, in this section are present the two first multiples of the mentioned primary reflections. [**CR**] ettore1/. const-off1,const-off

filters necessary for the calibration of the recorded vertical velocity. After performing the PZ summation step, having separated the pressure into up- and down-going energy, we improved the result of the up-going wavefield by employing an adaptive subtraction scheme. Our result demonstrates that the designed calibration filters were effective in decomposing almost all the events for all the offsets. In addition, the up-going wavefield obtained through the adaptive subtraction is less contaminated by the leakage of the down-going energy respect the same wavefield conducted using only the PZ summation algorithm. These two decomposed pressure fields enabled us to performing further processing, such as source signature estimation and debubbling.

ACKNOWLEDGMENTS

We thank Seabed Geosolutions for releasing the field data and for supporting the data handover by Paul Miligan and Shuki Ronen.

REFERENCES

Alvarez, G. and A. Guitton, 2007, Simultaneous adaptive matching of primaries and multiples with non-stationary filters: SEP-Report, **125**, 61–77.

Alves, G., 2014, Overview of the Moere Vest data set: SEP-Report, **155**.

Amundsen, L., 1993, Wavenumber-based filtering of marine point-source data: Geophysics, **58**, 1335–1348.

Barr, F. J. and J. L. Sanders, 1989, Attenuation of water-column reverations using pressure and velocity detectors in a water-bottom cable: SEG Expanded Abstracts, **59**, 653–656.

Grion, S., 2010, From PZ sum to wavefield separation, mirror imaging and up down deconvolution: The evolution of OBS data processing: KazGeo 2010, 83–87, European Association of Geoscientists and Engineers.

Grion, S., R. Exley, M. Manin, X. Miao, A. L. Pica, Y. Wang, P. Granger, and S. Ronen, 2007, Mirror imaging of OBS data: First Break, **25**, 37–42.

Melbø A. H. S., J. O. A. Robertsson, and D. Manen, 2002, PZ calibration by applying the equation of motion to critically refracted waves, 1030–1033. Soc. of Expl. Geophys.

Schalkwijk, K. M., C. P. A. Wapenaar, and D. J. Verschuur, 1999, Application of two-step decomposition to multicomponent ocean-bottom data: Theory and case study: Journal of Seismic Exploration, **8**, 261–278.

Sonneland, L., L. E. Berg, P. Eidsvig, A. Haugen, B. Fotland, and J. Vestby, 1986, 2-d deghosting using vertical receiver arrays: SEG Expanded Abstracts, **5**, 516.

Wong, M., B. Biondi, and S. Ronen, 2010, Joint least-squares inversion of up- and down-going signal for ocean bottom data sets: SEP-Report, **140**, 83–94.

Wong, M. and S. Ronen, 2009, Source signature and static shifts estimations for multi-component ocean bottom data: SEP-Report, **138**, 223–32.

Geophysical application of hyperbolic and hybrid L1/L2 optimization

Yinbin Ma, Musa Maharramov and Biondo Biondi

ABSTRACT

Hyperbolic penalty function method and hybrid L1/L2 methods, often generate better results than conventional least-squares solutions for inverse problem in geophysics. We apply a few hyperbolic and hybrid L1/L2 methods to 2-D Kirchhoff migration inversion and target-oriented linearized waveform inversion. The results demonstrate that we can recover a sparse/blocky model using hyperbolic and hybrid L1/L2 methods with acceptable computational cost.

INTRODUCTION

L1-norm optimization often yields more robost results compared with conventional least-squares optimization (Claerbout and Muir, 1973; Darche, 1989; Nichols, 1994; Guitton, 2005). Different algorithms for hyperbolic penalty function (HPF) optimization or L1/L2 hybrid optimization problems exist. All such algorithms require fine tuning of extra computational parameters. The computational cost and quality of results depends on the quality of parameter tuning. One concern is the physical interpretation of these parameters. Another concern is the cost of the solvers, as the extra computational cost associated with these techniques may not be justified by the improvement in the results. In this paper, we evaluate several hyperbolic and hybrid L1/L2 solvers.

We use several different solvers in our paper: least-squares with conjugate gradient (CGLS), iterative reweighted least squares (IRLS), Split-Bregman (Goldstein and Osher, 2009), Hyperbolic penalty function (HPF) with conjugate directions (Claerbout, 2009; Li et al., 2010; Zhang and Claerbout, 2010). We use IRLS and Split-Bregman methods when the objective function contain both L2-norm and L1-norm, and we call them L1/L2 hybrid methods. To characterize the performance, we first run a few iterations of least-squares methods, obtain the data residual and set it as threshold. Then we run L1 based techniques (HPF, IRLS and Split-Bregman) until the data residuals are below the threshold. The quality of inversion results and computational costs are then analyzed.

We test the solvers on several geophysical examples. First, we apply them to 2-D regularized Kirchhoff migration-inversion. The results show that hyperbolic and hybrid L1/L2 methods can recover the original model within acceptable amount of computation cost. Next, we apply the solvers to target-oriented linearized waveform inversion. We solve the problem described in Zhang and Claerbout (2010). Along with replicating the hyperbolic penalty function method result, we implement several other hybrid methods. While it is shown that hyperbolic and L1/L2 hybrid methods deliver significant improvement, we can not get pleasing results. We add a new L_1-based regularization term and produce better results.

METHOD

The general form of objective function we try to minimize has the following form:

$$\mathbf{J}(\mathbf{m}) = \frac{1}{2}\left\|\mathbf{Fm}-\mathbf{d}\right\|_2^2 + \frac{\varepsilon}{2}\left\|\mathbf{Am}\right\|_{norm}^{\alpha}. \tag{1}$$

The first term in the objective function is the data fitting term, and the second term is model regularization. Operator $\mathbf{A}$ can be an identity matrix if the model is sparse, and can be a total variation operator if the model is blocky. The second term is necessary in geophysical inverse problems, because we may have insufficient data. Even if we have sufficient data, we still want to keep the regularization term, because real data contains high amplitute burst noise, and overfitting the data may lead to inaccurate results in this case.

We use CGLS, IRLS, the Split-Bregman and Hyperbolic penalty function (HPF) with conjugate direction methods to solve the optimization problem with different norm (or penalty function).

We first use conventional least-squares solver CGLS. The regularization term is in L_2-norm, which means the objective takes the form,

$$\mathbf{J}(\mathbf{m}) = \frac{1}{2}\left\|\mathbf{Fm}-\mathbf{d}\right\|_2^2 + \frac{\varepsilon}{2}\left\|\mathbf{Am}\right\|_2^2. \tag{2}$$

Next, we implement a solver using hyperbolic penalty function (HPF), developed by Li et al. (2010) and Zhang and Claerbout (2010) . The solver solves

$$\mathbf{J}(\mathbf{m}) = \frac{1}{2}\left\|\mathbf{Fm}-\mathbf{d}\right\|_2^2 + \frac{\varepsilon}{2}\left\|\mathbf{Am}\right\|_{HB}, \tag{3}$$

where the second term is replaced by a nonlinear HPF $\|r\|_{HB} = \sqrt{r^2+R^2}-R$, where R is the threshold parameter. The penalty function approaches L1-norm as $r \gg R$, and approaches L2-norm as $r \ll R$. While using the HPF, the problem becomes nonlinear. Claerbout (2009) generalized the idea of the Conjugate Direction method for HPF optimization. At each step, we search the new direction in the plane spanned by the current gradient and the previous step. In this case, we need to solve the new direction by iterative methods. In the case of linear optimization, we can obtain the new direction directly. The detail of the method has been explained previously (Li et al., 2010; Zhang and Claerbout, 2010).

We also explore the behavior of solvers when the model styling term is L_1 norm as in equation (4). We use two different solvers, IRLS and the Split-Bregman method. Both methods are designed to solve the minimization problem of equation (4),

$$\mathbf{J}(\mathbf{m}) = \frac{1}{2}\left\|\mathbf{Fm}-\mathbf{d}\right\|_2^2 + \varepsilon\left\|\mathbf{Am}\right\|_1. \tag{4}$$

IRLS is an iterative method and at each step we solve a weighted least-squares. For the second term, the residual is now expressed as $\mathbf{r} = \mathbf{W_r Am}$. Where $diag(\mathbf{W_r})_i = |\mathbf{r_i}|^{-0.5}$. It can be found that $\|\mathbf{r}\|_2 = \|\mathbf{Am}\|_1$. Numerically, we minimize the following objective function using CGLS:

$$\mathbf{J}(\mathbf{m}) = \frac{1}{2}\left\|\mathbf{Fm}-\mathbf{d}\right\|_2^2 + \varepsilon\left\|\mathbf{W_r Am}\right\|_2^2. \tag{5}$$

After a few number of CG steps, we recalculate the weight function $\mathbf{W_r}$, and continued minimizing the new objective funcion. It is easy to see that, upon convergence, the $\mathbf{m}$ we obtained from equation (5) approaches the solution to equation (4).

We also implement an algorithm named the Split-Bregman method. The Split-Bregman method tries to minimize equation (4) by minimizing the following objective function:

$$\mathbf{J}(\mathbf{m},\mathbf{z}) = \frac{1}{2}\left\|\mathbf{Fm}-\mathbf{d}\right\|_2^2 + \frac{\rho}{2}\left\|\mathbf{Am}-\mathbf{z}\right\|_2^2 + \varepsilon\left\|\mathbf{z}\right\|_1. \tag{6}$$

With traditional optimization methods, we need to solve the problem with increasing ρ, and as $\rho \to \infty$ the solution converges to the original problem equation (4). Goldstein and Osher (2009) have designed the Split-Bregman method that could solve equation (6) with fixed ρ. They proved upon convergence we obtain the solution to equation (4). The Split-Bregman method is similar to alternating direction method of multipliers (ADMM) (Boyd et al., 2011).

To minimize equation (6) numerically, they introduce an additional parameter $\mathbf{b}$, and solve the following steps iteratively:

$$(\mathbf{m},\mathbf{z}) = \min_{\mathbf{m},\mathbf{z}}\{\frac{1}{2}\left\|\mathbf{Fm}-\mathbf{d}\right\|_2^2 + \lambda\left\|\mathbf{z}\right\|_1 + \frac{\rho}{2}\left\|\mathbf{z}-\mathbf{Am}-\mathbf{b}\right\|_2^2\}, \tag{7}$$

$$\mathbf{b} \leftarrow \mathbf{b} + (\mathbf{Am}-\mathbf{z}). \tag{8}$$

Where equation (7) are solved iteratively via:

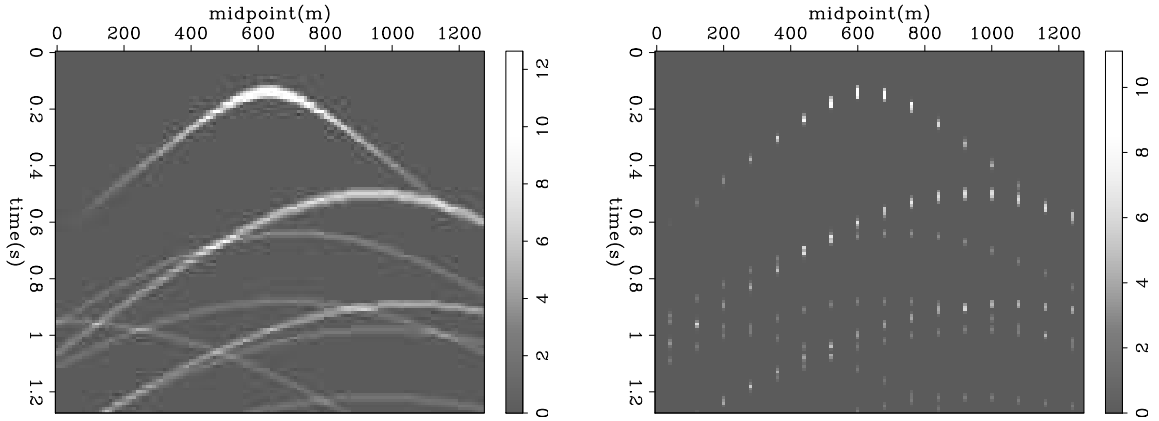
$$\mathbf{Step}\ 1 : \mathbf{m} = \min_{\mathbf{m}}\{\frac{1}{2}\left\|\mathbf{Fm}-\mathbf{d}\right\|_2^2 + \frac{\rho}{2}\left\|\mathbf{z}-\mathbf{Am}-\mathbf{b}\right\|_2^2\}, \tag{9}$$

$$\mathbf{Step}\ 2 : \mathbf{z} = \min_{\mathbf{m}}\{\lambda\left\|\mathbf{z}\right\|_1 + \frac{\rho}{2}\left\|\mathbf{z}-\mathbf{Am}-\mathbf{b}\right\|_2^2\}. \tag{10}$$

When the algorithm converges, from equation (8) we will have $\mathbf{Am}-\mathbf{z} \to \mathbf{0}$. The solution for the objective function in equation (6) converges to the solution for equation(4).

KIRCHHOFF MIGRATION INVERSION RESULTS

Kirchhoff migration involves the summation or spreading along travel time surfaces. It is well known that when we have insufficient data, the Kirchhoff migration does not yield good results because of data aliasing. To overcome this problem, we add geophysical constraints to our objective function and reduce the effect of aliasing. For simplicity, we use the Kirchhoff time migration as an example in this section, to test different solvers. The HPF



(a) (b)

Figure 1: Kirchhoff migration data. (a): with sufficient sampling; (b): with insufficient sampling. [**ER**] yinbin1/. kirchForward,kirchFSample

method has been tested on Kirchhoff migration in the previous work (Li et al., 2010). To be consistent, we use the same operator and same model. We formulate our problem as

$$\mathbf{Fm} \approx \mathbf{d} \tag{11}$$

$$\mathbf{Am} \approx \mathbf{0}, \tag{12}$$

to be consistent with the notation in the previous section. $\mathbf{F} = \mathbf{SK}$, where $\mathbf{K}$ is the Kirchhoff modeling operator, and $\mathbf{S}$ is the subsampling operator. $\mathbf{m}$ is reflectivity, and $\mathbf{d}$ is data recorded at the surface. $\mathbf{A}$ is regularization operator, and in the current example, $\mathbf{A} = \mathbf{I}$ as the model is reflectivity and it is sparse.

We choose the model space sampling to be 128×128, and subsampling data space to be 128×16, exactly the same as previous work (Li et al., 2010). As seen in Figure 1, the foward modeling operator maps the spikes in the model space to a series of hyperbolas in the data space if we have enough sampling points (Figure 1(a)), and it is aliased in our setting because we have insufficient data sampling (Figure 1(b)).

We solve the optimization problem with the four different methods mentioned in the previous section. First, we use the CGLS method for equation (2) with parameter $\varepsilon = 0.1$. We run the algorithm for 50 CG steps, and reduced the data residual from 60.93 to 0.33 ($\sim$ 0.54%). Next, we run HPF with equation (3), and successful reduce the data residual to the same level using 408 CG steps. Then, we run IRLS and Split-Bregman method for equation (4). We are able to match the data residual using 330 CG steps for IRLS and 417 CG steps for Split-Bregman. We can see the inversion results from different schemes in Figure 2, and the reconstructed data (Figure 3). The quality of inversion is greatly improved using hyperbolic and hybrid solvers. The additional computational cost is acceptable considering the improvement of inversion results.

We want to emphasis that we do not find optimized parameters for each solver. Thus, we cannot make a conclusion about comparison between different hyperbolic or hybrid solvers based on our numerical results.

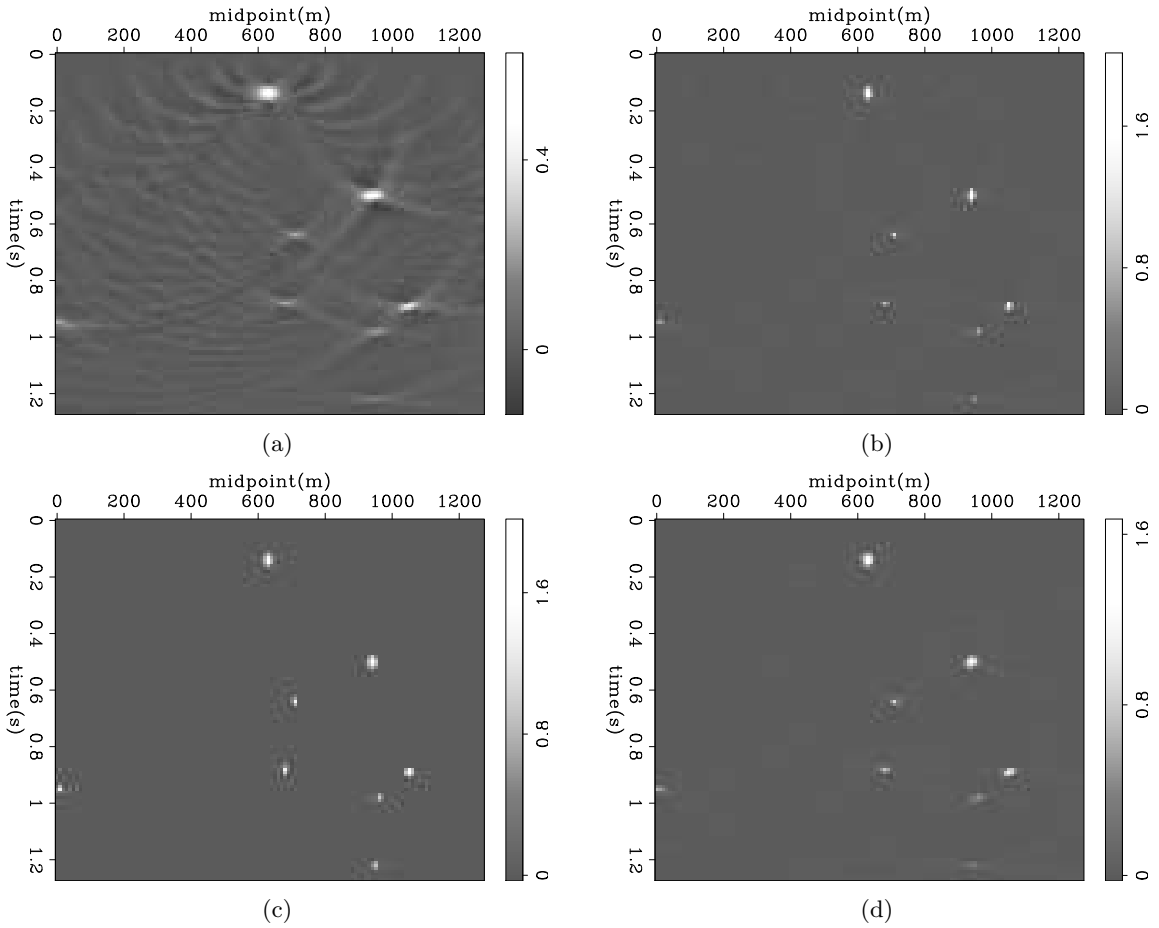


Figure 2: Comparison of the inversion results with different solvers: (a) CGLS, (b) HPF , (c) IRLS and (d) the Split-Bregman method. [**ER**] yinbin1/. kirchCGmod,kirchHBmod,kirchIRLSmod,kirchSPmod

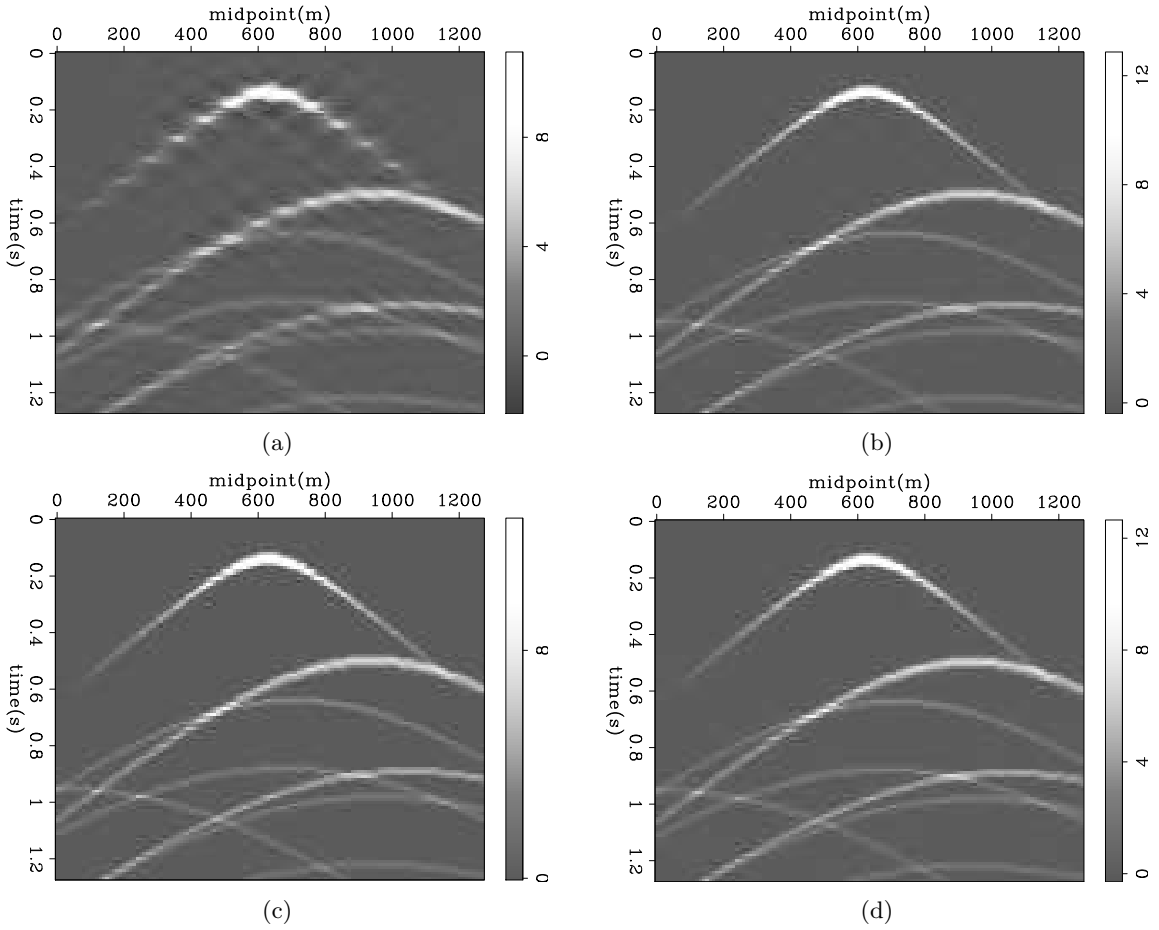


Figure 3: Comparison of modeled data for different schemes: (a) CGLS, (b) HPF, (c) IRLS and (d) the Split-Bregman method. [**ER**] yinbin1/. kirchCGmdata,kirchHBmdata,kirchIRLSmdata,kirchSPmdata

LINEARIZED WAVEFORM INVERSION RESULTS, PART 1

The target-oriented linearized waveform inversion has been explored in several previous works (M. Clapp and Biondi, 2005; Valenciano, 2006; Tang, 2008). The idea is from full waveform inversion(FWI) problem,

$$\mathbf{Lm} \approx \mathbf{d_{data}}. \tag{13}$$

We multiply $\mathbf{L}'$ on both side and obtain the normal equation as follows:

$$\mathbf{Hm} \approx \mathbf{m_{mig}} \tag{14}$$

where $\mathbf{H} = \mathbf{L}'\mathbf{L}$ is the Hessian operator, and $\mathbf{m_{mig}} = \mathbf{L}'\mathbf{d_{data}}$ is called migrated data.

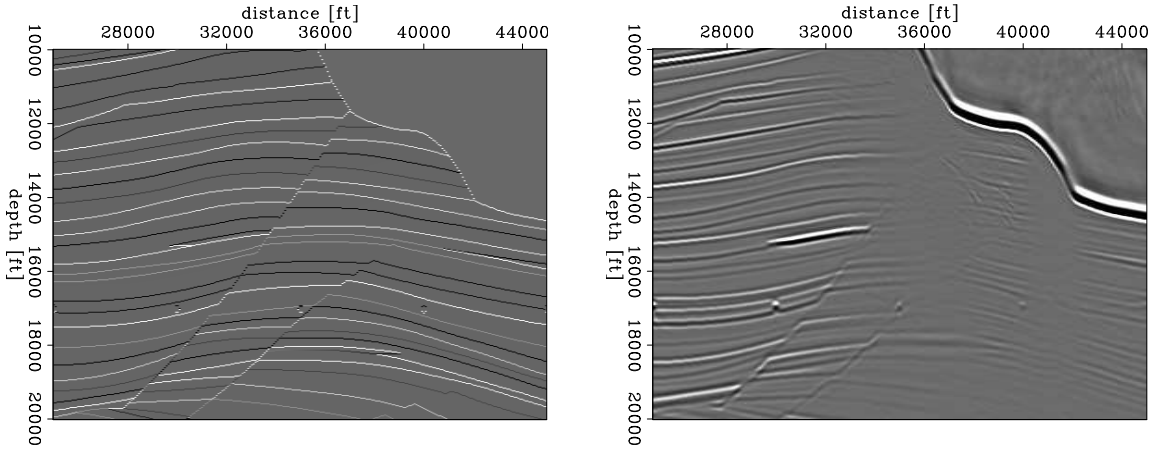


Figure 4: (a) Reflectivity model; (b) input migrated image. [**ER**] yinbin1/. refl-imaging,data-imaging-mig

The Hessian operator map vectors from the same model space, and we can view it as a convolution operator. It has been shown (Tang, 2008) that the Hessian operator is very sparse and we can afford to save the matrix. The advantage of equation (10) is that we can compute target-oriented inversion. However, under the salt body because of the complexity of wave propogation, the Hessian may be inaccurate or rank-deficient. It is challenging to obtain sparse and clear inversion results in the existence of salt body. As we can see in Figure 4(b), the migrated data has gaps that break the continuity of reflectors and the signal-to-noise ratio under the salt is small.

The previous work (Zhang and Claerbout, 2010) tried to solve the inversion problem with HPF as regularization. The results showd improvement comparing with least square method. In this part, we use $||\mathbf{m}||_{norm}$ as regularization, to be consistent with the previous results.

Similar to previous section, we solve the optimization problem with four different methods. The forward modeling operator is now Hessian $\mathbf{F} = \mathbf{H}$. And we choose the regularization operator to be identity $\mathbf{A} = \mathbf{I}$. The input data is $\mathbf{d} = \mathbf{m_{mig}}$.

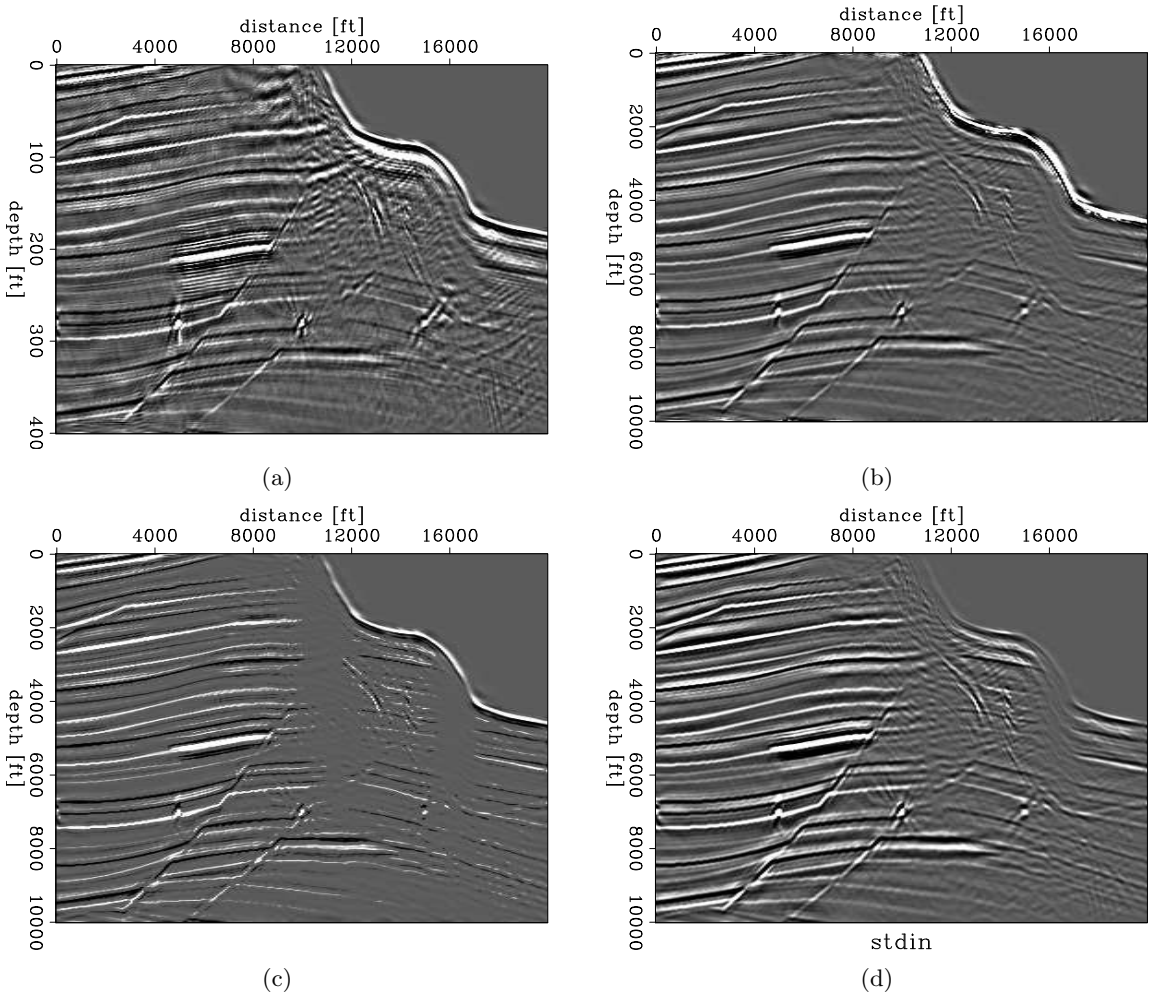


Figure 5: Comparison of linearized waveform inversion inversion results: (a) CGLS, (b) HPF, (c) IRLS and (d) the Split-Bregman method. [**ER**] yinbin1/. LSICGmod,LSIHBmod,LSIIRLSmod,LSISPmod

First, we use the CGLS method as the benchmark and run 50 CG steps. We reduce the data residual from 3.99 to 0.190($\sim$ 3.5%). Notice that for this problem, focusing on reducing the data residual is not a good choice. The data contains large error due to insufficient illumination under the salt, and the operator is not accurate. We search reasonable parameters for each method and ran less than 300 CG steps(3000 for HPF, 200 for IRLS, and 200 for Split-Bregman). The data residual is 0.199 for the HPF method, 0.311 for IRLS, and 0.254 for the Split-Bregman method.

All the hyperbolic and hybrid solvers improve the quality of the inversion results. However, none of them could yield satisfying results under the salt body. Figure 6 shows the data residual, suggesting the inversion results are not ideal.

LINEARIZED WAVEFORM INVERSION RESULTS PART 2

We obtain consistent results comparing with previous work. All of the solvers show similar behavior. We believe the poor quality of inversion results are not caused by the inappropriate use of the solvers. We need to check our objective function and make improvement in our model to obtain better results.

We use exactly the same regularization for Kirchhoff migration and linearized waveform inverion in the previous sections. It is obvious that in the Kirchhoff migration case, all we know is that the reflectors are a series of spikes and they are sparse. We used the correct regularization and obtained high-quality inversion results, as expected.

However, in the linearized waveform inversion problem, we know that the reflectors are spiky, and they are continuous along the $\mathbf{x}$ direction. Under the salt body, the continuity is more important than sparsity because of the small signal-to-noise ratio. In the previous section, we neglect part of the geophysical property, and should not expect to obtain high-quality results.

To make better use of our geophysical knowledge, we first use derivative operator on the x-axis as regularization,

$$\mathbf{J(m)} = \frac{1}{2}||\mathbf{Hm} - \mathbf{m_{mig}}||_2^2 + \varepsilon||\nabla_x \mathbf{m}||_1. \tag{15}$$

We can see significant improvement of the quality of inversion results from Figure 7(a). The reflectors under the salt body are cleaner. The gaps where we do not have enough data are filled. However,the reflectors are not smooth along x-direction near 17000 ft. It is not clear if they are improvements or artifacts.

Next, we put both sparsity and continuity information into the objective function as follows:

$$\mathbf{J(m)} = \frac{1}{2}||\mathbf{Hm} - \mathbf{m_{mig}}||_2^2 + \varepsilon_\mathbf{1}||\mathbf{m}||_1 + \varepsilon_\mathbf{2}||\nabla_x \mathbf{m}||_1. \tag{16}$$

As the wave propogation is complicated because of the existence of salt body, the data under the salt is diminished and twisted. It makes sense to choose $\varepsilon_\mathbf{2} >> \varepsilon_\mathbf{1}$ to recover the feature under the salt. Figure 8 shows the inversion result with the new objective function. We can see the reflectors under the salt body are recovered.

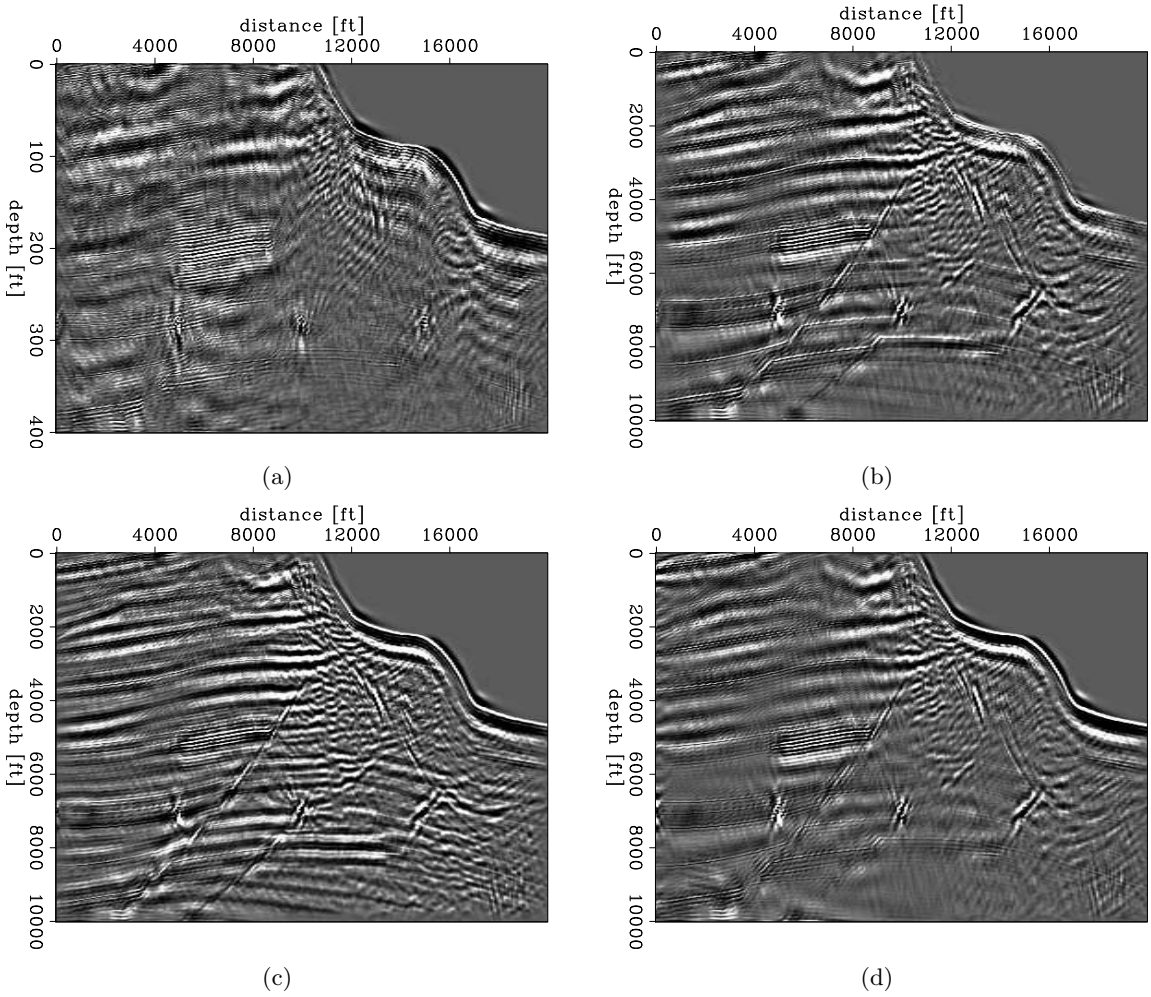


Figure 6: Comparison of linearized waveform inversion data residual: (a) CGLS, (b) HPF, (c) IRLS and (d) the Split-Bregman method. All figures are clipped at the same level. [**ER**] yinbin1/. LSICGresd,LSIHBresd,LSIIRLSresd,LSISPresd

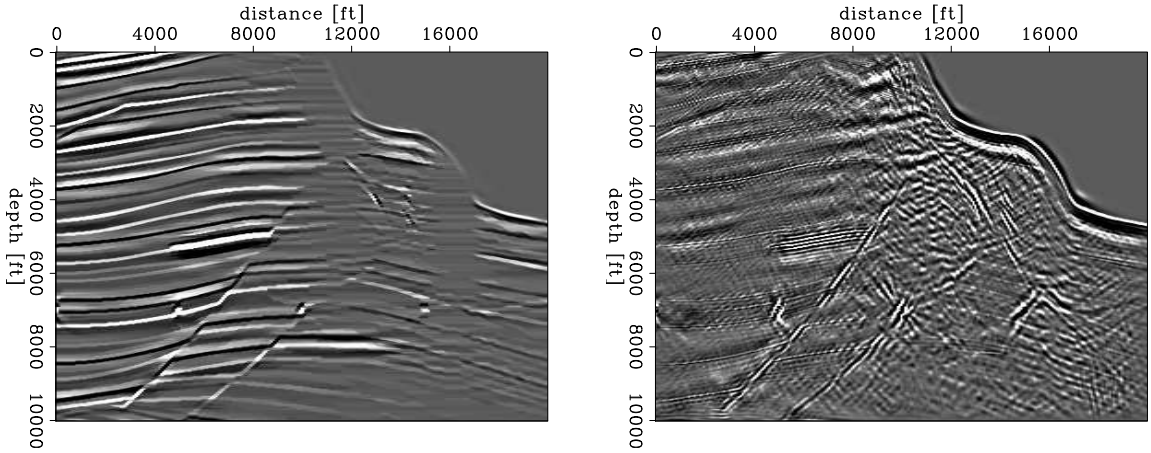


Figure 7: Linearized waveform inversion result with derivative regularization terms using IRLS: (a) inversion result, (b) data residual. [**ER**] yinbin1/. LSIIRLSderivmod,LSIIRLSderivresd

CONCLUSIONS

Hyperbolic and hybrid L1/L2 solvers generate more sparse and blocky models, they are also more expensive than traditional least-squares methods. We check a couple of solvers on simple geophysical problems. We demonstrate that great improvement of inversion quality is possible, with acceptable computational cost. However, it is still challenging to find a good L1 (L1-type) solver for generic geophysical problems. In the future, we need to explore the linearized waveform inversion model in detail to get a better understanding. We also need to test a more sophisticated model to get a better understanding of the solvers.

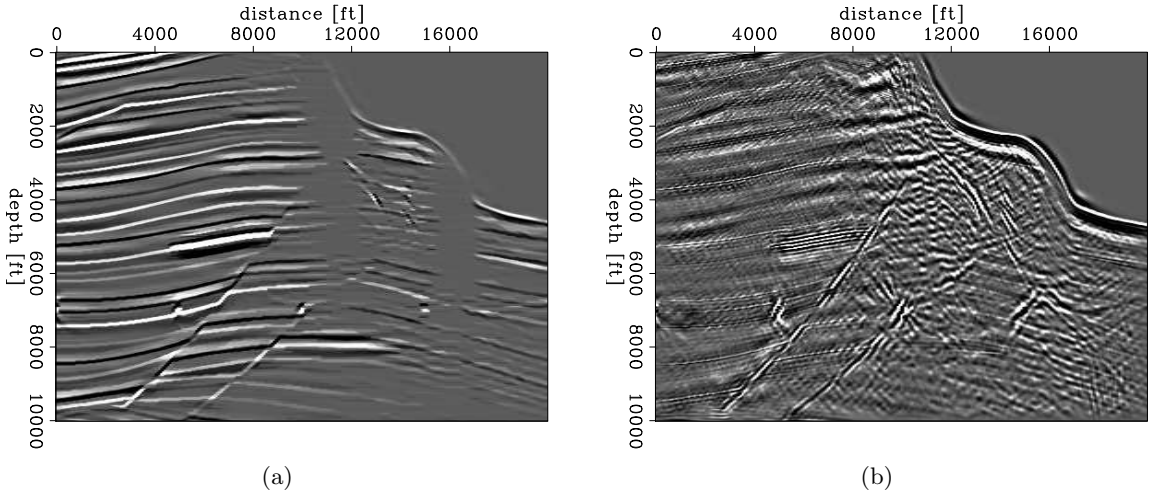


Figure 8: Linearized waveform inversion result with both regularization terms. Using IRLS. (a) inversion result, (b) data residual. [**ER**] yinbin1/. LSIIRLStotmod,LSIIRLStotresd

ACKNOWLEDGEMENT

We thank Elita Li, Mandy Wong and Jon Claerbout for useful suggestions. We also would like to thank Robert Clapp for providing numerious ideas and techinical support on implementing the solvers in the new SEP C++ library.

REFERENCES

Boyd, S., N. Parikh, E. Chu, B. Peleato, and J. Eckstein, 2011, Distributed optimization and statistical learning via the alternating direction method of multipliers: Found. Trends Mach. Learn., **3**, 1–122.

Claerbout, J. F., 2009, Blocky models via the L1/L2 hybrid norm: SEP-Report, **139**, 1–10.

Claerbout, J. F. and F. Muir, 1973, Robust modeling with erratic data: Geophysics, **18**, 826–844.

Darche, G., 1989, Iterative L_1 deconvolution: SEP 61, 281–302.

Goldstein, T. and S. Osher, 2009, The split bregman method for l1-regularized problems: SIAM Journal on Imaging Sciences, **2**, 323–343.

Guitton, A., 2005, Multidimensional seismic noise attenuation: PhD thesis, Stanford University.

Li, E., Y. Zhang, and J. F. Claerbout, 2010, Geophysical applications of a novel and robust L1 solver: SEP 140, 155–164.

M. Clapp, R. C. and B. Biondi, 2005, Regularized least-squares inversion for 3-D subsalt imaging: Soc. of Expl. Geophys., 1814–1817.

Nichols, D., 1994, Velocity-stack inversion using L_p norms: SEP 82, 1–16.

Tang, Y., 2008, Wave-equation hessian by phase encoding: SEP-Report, **134**, 1–25.

Valenciano, A., 2006, Target-oriented wave-equation inversion: Geophysics, **71**, A35–38.

Zhang, Y. and J. F. Claerbout, 2010, Least-squares imaging and deconvolution using the hybrid norm conjugate-direction solver: SEP 140, 129–142.

Synthetic model building using a simplified basin modeling approach

Robert G. Clapp

ABSTRACT

Generating a realistic synthetic model is a challenging problem in a geophysical research environment. Achieving the right balance between being complex enough to be realistic while still simple enough that a new algorithm can be debugged is hard to achieve. I propose a different way to generate synthetic models, by allowing the user to specify a series of geologic events. The result of each event is approximated on the current model. This approach has the benefit of allowing complex models to be built is easily extendable to multiple model parameters, and allows the user to "turn off" events, allowing the construction of simpler models by stages. A geologic event-based modeling strategy proves to be useful to build simple to quite complex models.

INTRODUCTION

Synthetic models play an important role in geophysical research. Over the last 25 years, SEG and company generated synthetics (Versteeg, 1993) have been used extensively by researchers throughout the world. These synthetics have found such wide use for two reasons. First, generating realistic synthetic models is difficult, particularly 3-D models. Tools such as Gocad (Cain et al., 1998) allow the user to specify layers, faults, and other geologic features before creating a 3-D grid to ease the process, but still require significant experience to use effectively. Second, industry wide synthetics provide a common benchmark to compare results from different companies. These models are useful for final testing of a finished algorithm, but often are too complex to be used in algorithmic development.

Another approach to generating synthetic models is following the methodology of basin modelers. Basin modeling takes a more geologic approach to describing models than the more computer science-based approach followed by GoCad and its competitors. Basin modeling attempts to model the geologic history of a given piece of the earth to explain its current properties.

In this paper I build on the tool introduced in Clapp (2013) for generating more realistic synthetic models. Specifically I improve the fault, emplacement, and river erosion models while adding unconformities and compressional folding. In this paper I describe these new modules, introduce a python script that simplifies the model building process, and provide a complete list of all of the parameters for each model building module.

FAULTS

The simplified view of faulting that you see in introductory geology textbooks where you a see bulk shift of all the layers on one side of the fault move up, or down, or side to side,

does not accurately reflect how faulting actually occurs at the large scale. Movement away from the fault location decays away from fault surface in all directions. In addition fault planes are not the planar surfaces that one often sees depicted in illustrations.

In `ModelCreate` faults are created on a cylndrical coordinate system. The user describes the beginning and ending of the fault plane in terms of a rectangular coordinate system. These are then mapped into a polar coordinate system by defining the angular range represented by the beginning and ending points of the fault surface. Using this methodology, a user can make a fault virtually planar by defining a small angular range and curved using a larger angular range. The user defines the amount of movement in terms of angular shift. Figure 1 graphically illustrates how the fault is described.

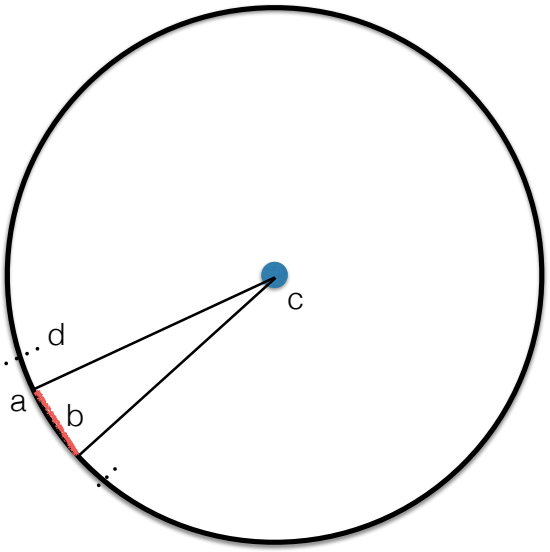


Figure 1: Fault movement is defined by an initial point 'A' and a point 'b' defined by distance away from 'a' in z and x. In addition the user defines the arc length separating points a and b (the red line). Using the arc length along with points 'a' and 'b' a circle is implied with center 'c'. Fault movement is then described in terms of this circle. The user specifies the die out of fault movement 'd' in terms of arc length along with the shift in radians along the circle. bob1/. circle

The decline in movement along the fault is also defined in terms of the cylindrical coordinate system in terms of radial distance, angular range, and perpindicular distance. This coordinate system can then be rotated to an arbitrary azimuth. Figure 2 shows two faults. The fault on the left shows a nearly linear fault, using a small angular range. The fault on the right shows a more curved fault surface using a large angular range.

RIVER CHANELS

The previous version of `ModelCreate` (Clapp, 2013) was capable of producing river channels but required significant parameterization by the user to get reasonable shapes. As a result I overhauled the module and added additional complexity. The user now specifies a general direction for the river in terms of azimuth. The user also specifies a wavelength and wave amplitude which gives the river its sinusoidal like appearance In addition the user specifies a randomness factor. This factor is used to add semi-random other wavelengths to the river's path, creating a more realistic pattern. In addition the user can specify a number of levels in the river's paths. This parameter takes into account the slope of the river's path to create deeper river bed sediments that are consistent with the river's curvature. Figure 3 shows an example of using these features. The left panel shows several different rivers. The right panel shows a smaller region of the model where deeper river channels can easily be seen.

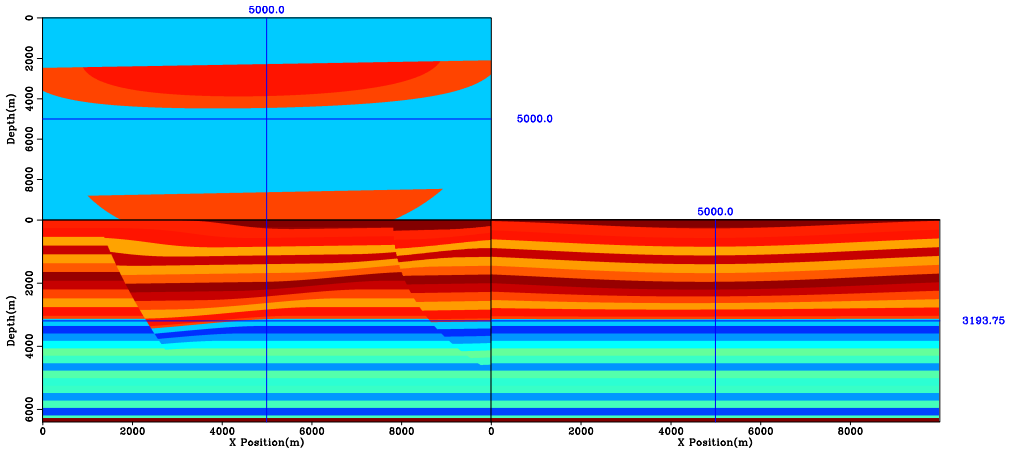


Figure 2: Two examples of faults. In the lower left fault a distant center point is calculated by specifying a small angular range compared to shift in rectangular coordinated. The upper right fault is curved by specifying a large angle compare rectangular distance resulting in a smaller circle (see explanation in Figure 1. bob1/. fault

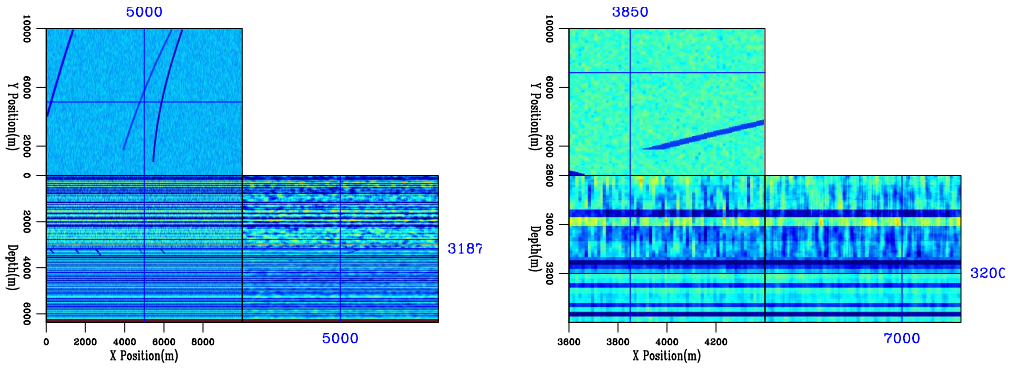


Figure 3: The left panel shows an example of constructing several different river channels. The right panels shows a blowup around the river chanels. Note the stacked river channel feature where the river channel has moved over time. bob1/. river

EMPLACEMENT

Describing salt flow, which would follows the spirit of `ModelCreate` , is quite challenging and is not something I have undertaken at this time. Instead I allow bodies to be emplaced into the geologic scenario. Just replacing one earth property with another does not produce a pleasing model in all cases, because sediments are affected by the flowing in of salt. As a result the emplacement module offers the ability to warp the surrounding structure. To do this I first smooth the emplaced body. This creates a field that is at its maximum inside the body and dies out away from the body. This field is used the shift sediments up. Shifts are greatest close to the salt body and die off as you get further from body. Note how the sediments are drawn up to the sides along the salt and the structure of the salt top and bottom can be seen around the salt in Figure 4. In addition I allow the user to change the earth's properties under the salt (e.g. underpressure). Again this is a smooth field that dies off away from the salt.

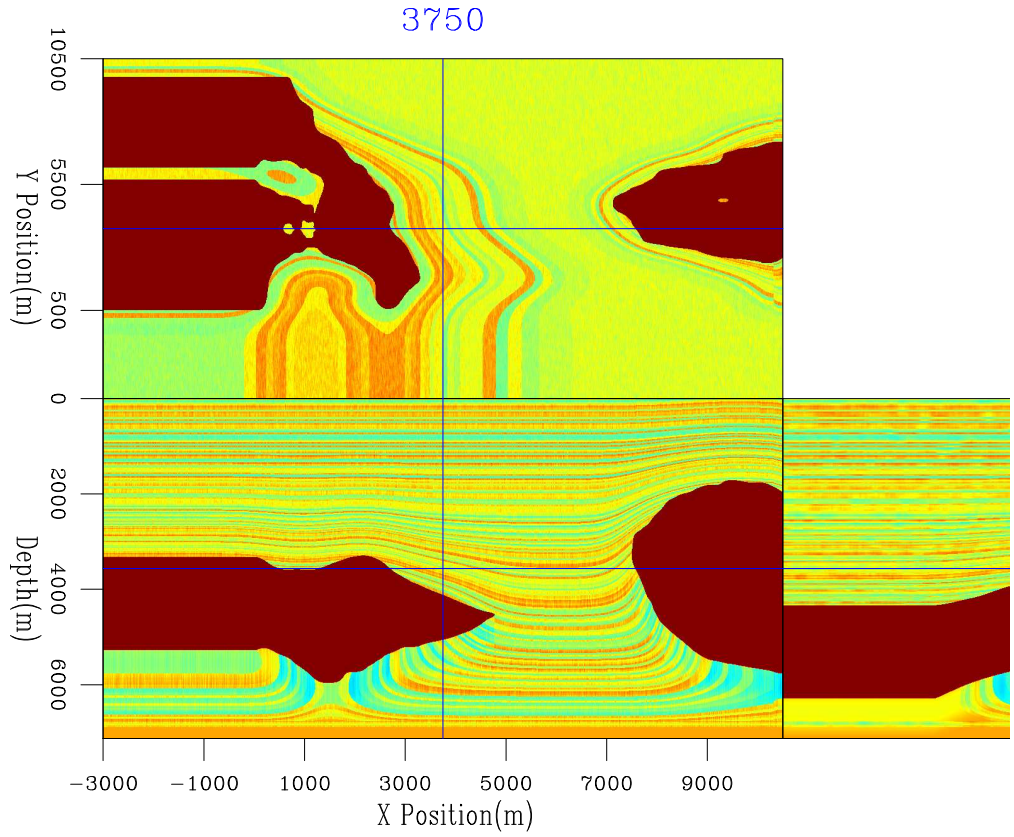


Figure 4: An example of emplacing a salt body into a model. In this case the sediments around the salt have been modified so that they are affected by the emplacement. In addition a low velocity region has been added below the salt. bob1/. implace

FOLDING AND EROSION

In the original version of `ModelCreate` I allowed the user to create hills. Geologically, hills alone are usually associated with some compressional event that also leads to valleys,

and there is often some level of spatial consistency because the entire region is under the same compessional regime. As a result I added a module that allows the user to describe a compressional stress direction and the amount of uplift associated with it. These two values, along with the average wavelength of the hill/valley pattern and a description of randomness inline and crossline are used to calculate a spatial field of vertical shifts up and down. Figure 5 shows an example of using this geomodule. Note how the hill/valley pattern, while showing a dominant wavelength, also has spatial variation.

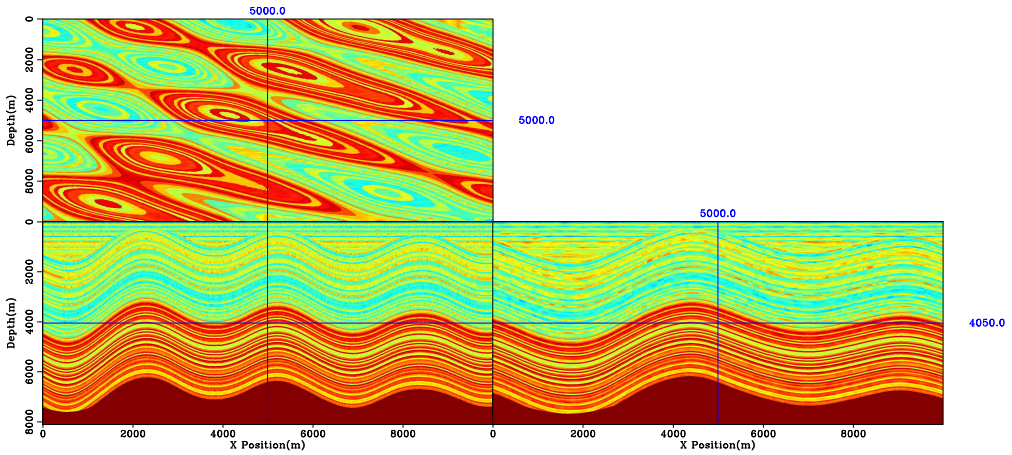


Figure 5: An example of introducing compressionalforces and folding the layers. Note how the folding pattern is somewhat repeatable but still varies significantly. In addition, but harder to see, the thickness of the various beds change spatially. bob1/. squish

I also added a third erosion option (to go along with a river channel and bowls), a planar erosion event. With the unconformity option, all of the model down to a certain depth is scraped off, to be replaced with later sediment deposits. Figure 6 shows the result of eroding the hill/valley pattern off to certain depth (note how some of anticlines are cut) and then adding additional layering over it.

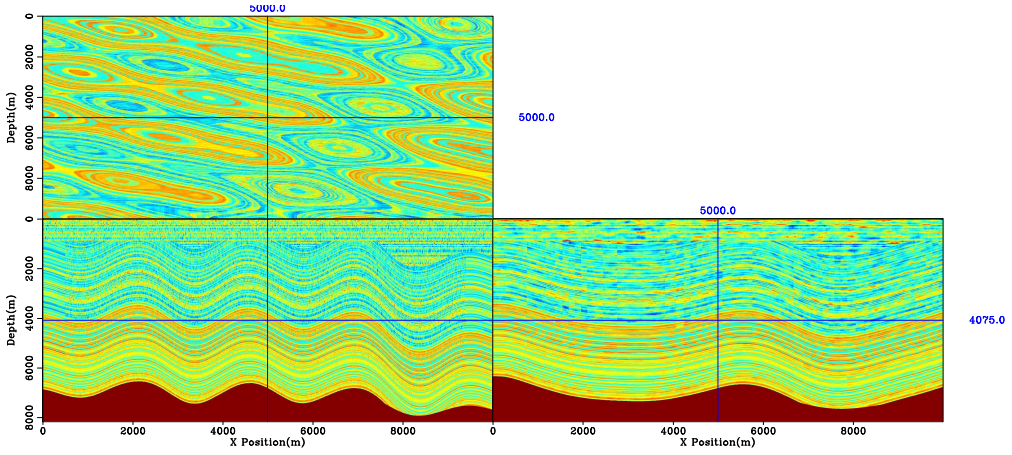


Figure 6: An example of a horizontal unconformity event. Note how the top of some of the anticlinal structures has been eroded off. bob1/. unconformity

GEOMODEL.PY

Creating a realistic geologic model using `ModelCreate` can involve hundreds of different geologic events. Given that each event has up to 12 parameters it can quickly become a daunting task. To make it a more reasonable task, `Geomodel` is a python library to ease the model building process.

The basic idea of the library is that specifying every parameter to every module at every step is too time consuming. A better idea is to specify a reasonable range of possible parameters for each geologic event (controlled by a `geomodule`). When the user asks for an event to be added to the geologic history, the parameters are randomly chosen from the acceptable range the user specified. The user at anytime has the option of changing the possible parameters range.

For example, to create the fault model shown in Figure 2 I began by importing the `Geomodel` library and making a copy of the library default parameter set.

```
import Geomodel
myDefaults=Geomodel.defaults()
```

I then changed some of the default parameters for the deposit and fault modules.

```
myDefaults.change_param_ranges("deposit",
 ["thick:500:500","dev_pos:.2:.4","dev_layer:.2:.29","layer:11:39"])
myDefaults.change_param_ranges("fault",["begz:.1:.3","begy:.1:.9",
 "perp_die:.5:.9","dist_die:1.2:16","theta_shift:1.:1.2"])
```

I specified that I wanted to create a new model using my default parameter set, the sampling in z, the axes in x and y, and the parameter file I wanted to write out to.

```
mod=Geomodel.model(myDefaults,6.25,800,0,12.5,800,0,12.5,3000,20,"temp.P")
```

I then changed the velocity range I wanted for my first layer and added it to the model.

```
mod.change_param_ranges("deposit",["prop:2200:2600"])
mod.add_event("deposit")
```

I created a new layer, changed the velocity range, and deposited another layer.

```
mod.new_layer()
mod.change_param_ranges("deposit",["prop:2300:2300"])
mod.add_event("deposit")
```

I next change some of the parameters for the fault model, created a fault event, changed the parameters again, and created another fault event.

```
mod.change_param_ranges("fault",["begx:.1:.4","dz:1000:2000","daz:500:502",
```

```
  "azimuth:1:3","deltaTheta:4:5","theta_die:2:4"])
mod.add_event("fault")
mod.change_param_ranges("fault",["begx:.55:.9","dz:2000:4000","daz:500:502",
"azimuth:1:3","deltaTheta:32:36","theta_die:20:23"])
mod.add_event("fault")
```

Finally, I finished my geologic model and wrote its description to the parameter file.

```
mod.finish()
```

Using this scheme, complex models can be created. Figure 7 shows three slices through a 3-D model containing six layers, eighteen faults, two compressional events, fifteen gaussian anomalies, and five river channels.

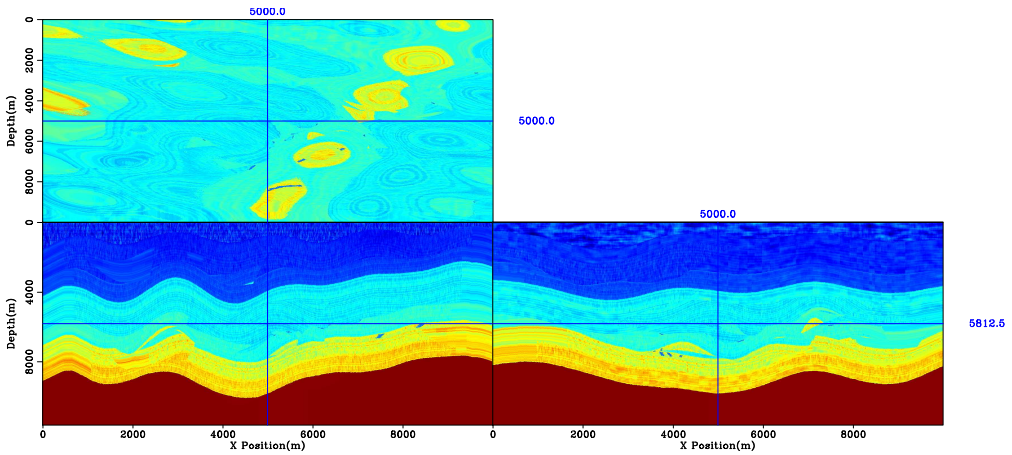


Figure 7: Three views of a complex synthetic containing deposition, river channels, faults, compressional folding, and gaussian anomalies. bob1/. big

CONCLUSIONS

Complex synthetic models are useful not only for benchmarking but in the code development process. Complex models can be built up by simulating a series of geologic events such as deposition, erosion, compressional events, and faulting. `ModelCreate` and `Geomodel` used together allow for relatively easy construction of complex models.

REFERENCES

Cain, G., G. Cambois, M. Gehin, and R. Hall, 1998, Reducing risk in seismic acquisition and interpretation of complex targets using a Gocad-based 3-D modeling tool: 68th Ann. Internat. Mtg, 2072–2075, Soc. of Expl. Geophys.

Clapp, R. G., 2013, Synthetic model building using a simplified basin modeling approach: SEP-Report, **149**, 273–280.

Versteeg, R. J., 1993, Sensitivity of prestack depth migration to the velocity model: Geophysics, **58**, 873–882.

Elastic modeling of surface waves

Gustavo Alves

ABSTRACT

I implement a free surface boundary condition for the generation of surface waves using a 10^{th} order in space and 2^{nd} order in time finite-difference staggered-grid scheme. I show an example of a field seismic section and recreate its main features using the proposed scheme. The synthetic data created show Rayleigh waves, backscattered waves and mode conversions, and fit the kinematics of the field data.

INTRODUCTION

Surface waves appear whenever the interface between two elastic media can be described as a free surface, i.e., a boundary that exhibits null stress components acting on the plane of the interface (Takeuchi and Saito, 1972). In seismology, this condition is observed in ground-to-air and sea floor interfaces. Therefore, surface waves are ubiquitous in seismic land data and increasingly more common in marine data with the advent of sea bottom receivers like ocean bottom cables (OBCs) and ocean bottom nodes (OBNs) (Boiero et al., 2013).

In seismic data, surface waves are observed as slow, dispersive linear high wavenumber events and are usually considered coherent noise that needs to be either filtered or muted out (Boustani et al., 2013). However, they contain important information about the elastic properties of the shallow layers of the subsurface and could give hints into their understanding. This makes surface wave modeling an interesting and yet understudied subject in seismic.

In Alves and Biondi (2014), I focused in the implementation of a finite-difference staggered-grid method for modeling elastic waves. That work was based on the work of Virieux (1986) and followed the algorithm later proposed by Ikelle and Amundsen (2005). Here, I extend the previous work, adding a free-surface boundary condition that allows the generation of surface waves. The methodology still follows that presented in Ikelle and Amundsen (2005), but the spatial derivative stencils used are 10^{th} order. In the next section, I present the changes to the implementation adopted to achieve this higher order boundary condition.

Finally, I compare a field seismic section and an equivalent synthetic data set generated using the proposed algorithm.

METHOD

The introduction of a free surface in the elastic finite-difference code imposes a non-slipping contact and null stresses at the boundary, thus satisfying the continuity relations for strain and stress. I implement these boundary conditions in a 10^{th} order stencil using the method of

mirror images, similarly to the 4^{th} order implementation described in Ikelle and Amundsen (2005). However, due to the longer stencil, the method of mirror images must be applied gradually as the differential equation is evaluated at different distances from the free surface. Equations 1 through 4 describe the normal stress calculated at the grid points close to the boundary at a fixed time step. For the solution shown here, the free surface is located at $z = 1$ and the indeces correspond to the grid positions where the properties are evaluated:

$$\begin{aligned}
\tau_{zz}(x,1) &= \frac{\Delta t}{\Delta x}((\lambda(x,1)+2\mu(x,1))\times \\
&(\quad c_1 \quad (V_z(x,6)+V_z(x,4)) \\
&+ \quad c_2 \quad (V_z(x,5)+V_z(x,3)) \\
&+ \quad c_3 \quad (V_z(x,4)+V_z(x,2)) \\
&+ \quad c_4 \quad (V_z(x,3)+V_z(x,1)) \\
&+ \quad c_5 \quad (V_z(x,2)-V_z(x,1))) \\
&+ \quad \lambda(x,1)\times \\
&(\quad c_1 \quad (V_x(x+5,1)-V_x(x-4,1)) \\
&+ \quad c_2 \quad (V_x(x+4,1)-V_x(x-3,1)) \\
&+ \quad c_3 \quad (V_x(x+3,1)-V_x(x-2,1)) \\
&+ \quad c_4 \quad (V_x(x+2,1)-V_x(x-1,1)) \\
&+ \quad c_5 \quad (V_x(x+1,1)-V_x(x,1)))); \qquad (1)
\end{aligned}$$

$$\begin{aligned}
\tau_{zz}(x,2) &= \frac{\Delta t}{\Delta x}((\lambda(x,2)+2\mu(x,2))\times \\
&(\quad c_1 \quad (V_z(x,7)+V_z(x,3)) \\
&+ \quad c_2 \quad (V_z(x,6)+V_z(x,2)) \\
&+ \quad c_3 \quad (V_z(x,5)+V_z(x,1)) \\
&+ \quad c_4 \quad (V_z(x,4)-V_z(x,1)) \\
&+ \quad c_5 \quad (V_z(x,3)-V_z(x,2))) \\
&+ \quad \lambda(x,2)\times \\
&(\quad c_1 \quad (V_x(x+5,1)-V_x(x-4,1)) \\
&+ \quad c_2 \quad (V_x(x+4,1)-V_x(x-3,1)) \\
&+ \quad c_3 \quad (V_x(x+3,1)-V_x(x-2,1)) \\
&+ \quad c_4 \quad (V_x(x+2,1)-V_x(x-1,1)) \\
&+ \quad c_5 \quad (V_x(x+1,1)-V_x(x,1)))); \qquad (2)
\end{aligned}$$

$$
\begin{aligned}
\tau_{zz}(x,3) &= \frac{\Delta t}{\Delta x}((\lambda(x,3)+2\mu(x,3))\times \\
&(\quad c_1 \quad (V_z(x,8)+V_z(x,2)) \\
&+ \quad c_2 \quad (V_z(x,7)+V_z(x,1)) \\
&+ \quad c_3 \quad (V_z(x,6)-V_z(x,1)) \\
&+ \quad c_4 \quad (V_z(x,5)-V_z(x,2)) \\
&+ \quad c_5 \quad (V_z(x,4)-V_z(x,3))) \\
&+ \qquad \lambda(x,3)\times \\
&(\quad c_1 \quad (V_x(x+5,1)-V_x(x-4,1)) \\
&+ \quad c_2 \quad (V_x(x+4,1)-V_x(x-3,1)) \\
&+ \quad c_3 \quad (V_x(x+3,1)-V_x(x-2,1)) \\
&+ \quad c_4 \quad (V_x(x+2,1)-V_x(x-1,1)) \\
&+ \quad c_5 \quad (V_x(x+1,1)-V_x(x,1)));
\end{aligned}
\tag{3}
$$

$$
\begin{aligned}
\tau_{zz}(x,4) &= \frac{\Delta t}{\Delta x}((\lambda(x,4)+2\mu(x,4))\times \\
&(\quad c_1 \quad (V_z(x,9)+V_z(x,1)) \\
&+ \quad c_2 \quad (V_z(x,8)-V_z(x,1)) \\
&+ \quad c_3 \quad (V_z(x,7)-V_z(x,2)) \\
&+ \quad c_4 \quad (V_z(x,6)-V_z(x,3)) \\
&+ \quad c_5 \quad (V_z(x,5)-V_z(x,4))) \\
&+ \qquad \lambda(x,3)\times \\
&(\quad c_1 \quad (V_x(x+5,1)-V_x(x-4,1)) \\
&+ \quad c_2 \quad (V_x(x+4,1)-V_x(x-3,1)) \\
&+ \quad c_3 \quad (V_x(x+3,1)-V_x(x-2,1)) \\
&+ \quad c_4 \quad (V_x(x+2,1)-V_x(x-1,1)) \\
&+ \quad c_5 \quad (V_x(x+1,1)-V_x(x,1)))),
\end{aligned}
\tag{4}
$$

where $\tau_{zz}(x,z)$ is the normal stress component in the z direction; V_x and V_z are the particle displacements in the x and z directions, respectively; and $\lambda(x,z)$ and $\mu(x,z)$ are the Lamé parameters. Equation 5 shows the values of c_1 to c_5, which correspond to the 10^{th} order finite-difference coefficients, according to Liu and Sen (2009).

$$
\begin{aligned}
c_1 &= \frac{35}{294912}, \\
c_2 &= -\frac{405}{229376}, \\
c_3 &= \frac{567}{40960}, \\
c_4 &= -\frac{735}{8192}, \\
c_5 &= \frac{19845}{16384}.
\end{aligned}
\tag{5}
$$

The equations for the horizontal stress can be obtained by swapping the $(\lambda(x,z) + 2\mu(x,z))$ and $\lambda(x,z)$ terms in Equations 1 through 4. For the shear stress τ_{xz}, the derivatives of the displacement field are shown on Equations 6 through 10. Continuity of the physical parameters requires the shear stress be null exactly at the boundary at $z = 1$. The coefficients for the 10^{th} order stencil are the same as in the previous set of equations.

$$\tau_{xz}(x,1) = 0 \tag{6}$$

$$\begin{aligned}
\tau_{xz}(x,2) &= \frac{\Delta t}{\Delta x}((\mu(x,2)) \times \\
&(\quad c_1 \quad (V_x(x,6) + V_x(x,4)) \\
&+ \quad c_2 \quad (V_x(x,5) + V_x(x,3)) \\
&+ \quad c_3 \quad (V_x(x,4) + V_x(x,2)) \\
&+ \quad c_4 \quad (V_x(x,3) + V_x(x,1)) \\
&+ \quad c_5 \quad (V_x(x,2) - V_x(x,1)) \\
&+ \quad c_1 \quad (V_z(x+4,2) - V_z(x-5,2)) \\
&+ \quad c_2 \quad (V_z(x+3,2) - V_z(x-4,2)) \\
&+ \quad c_3 \quad (V_z(x+2,2) - V_z(x-3,2)) \\
&+ \quad c_4 \quad (V_z(x+1,2) - V_z(x-2,2)) \\
&+ \quad c_5 \quad (V_z(x,2) - V_z(x-1,2)))),
\end{aligned} \tag{7}$$

$$\begin{aligned}
\tau_{xz}(x,3) &= \frac{\Delta t}{\Delta x}((\mu(x,3)) \times \\
&(\quad c_1 \quad (V_x(x,7) + V_x(x,3)) \\
&+ \quad c_2 \quad (V_x(x,6) + V_x(x,2)) \\
&+ \quad c_3 \quad (V_x(x,5) + V_x(x,1)) \\
&+ \quad c_4 \quad (V_x(x,4) - V_x(x,1)) \\
&+ \quad c_5 \quad (V_x(x,3) - V_x(x,2)) \\
&+ \quad c_1 \quad (V_z(x+4,3) - V_z(x-5,3)) \\
&+ \quad c_2 \quad (V_z(x+3,3) - V_z(x-4,3)) \\
&+ \quad c_3 \quad (V_z(x+2,3) - V_z(x-3,3)) \\
&+ \quad c_4 \quad (V_z(x+1,3) - V_z(x-2,3)) \\
&+ \quad c_5 \quad (V_z(x,3) - V_z(x-1,3)))),
\end{aligned} \tag{8}$$

$$
\begin{aligned}
\tau_{xz}(x,4) &= \frac{\Delta t}{\Delta x}((\mu(x,4)) \times \\
&(\quad c_1 \quad (V_x(x,8) + V_x(x,2)) \\
&+ \quad c_2 \quad (V_x(x,7) + V_x(x,1)) \\
&+ \quad c_3 \quad (V_x(x,6) - V_x(x,1)) \\
&+ \quad c_4 \quad (V_x(x,5) - V_x(x,2)) \\
&+ \quad c_5 \quad (V_x(x,4) - V_x(x,3)) \\
&+ \quad c_1 \quad (V_z(x+4,4) - V_z(x-5,4)) \\
&+ \quad c_2 \quad (V_z(x+3,4) - V_z(x-4,4)) \\
&+ \quad c_3 \quad (V_z(x+2,4) - V_z(x-3,4)) \\
&+ \quad c_4 \quad (V_z(x+1,4) - V_z(x-2,4)) \\
&+ \quad c_5 \quad (V_z(x,4) - V_z(x-1,4)))),
\end{aligned}
\tag{9}
$$

$$
\begin{aligned}
\tau_{xz}(x,5) &= \frac{\Delta t}{\Delta x}((\mu(x,5)) \times \\
&(\quad c_1 \quad (V_x(x,9) + V_x(x,1)) \\
&+ \quad c_2 \quad (V_x(x,8) - V_x(x,1)) \\
&+ \quad c_3 \quad (V_x(x,7) - V_x(x,2)) \\
&+ \quad c_4 \quad (V_x(x,6) - V_x(x,3)) \\
&+ \quad c_5 \quad (V_x(x,5) - V_x(x,4)) \\
&+ \quad c_1 \quad (V_z(x+4,5) - V_z(x-5,5)) \\
&+ \quad c_2 \quad (V_z(x+3,5) - V_z(x-4,5)) \\
&+ \quad c_3 \quad (V_z(x+2,5) - V_z(x-3,5)) \\
&+ \quad c_4 \quad (V_z(x+1,5) - V_z(x-2,5)) \\
&+ \quad c_5 \quad (V_z(x,5) - V_z(x-1,5)))).
\end{aligned}
\tag{10}
$$

It is important to note that the spatial derivatives calculated from the displacements are not centered at the same indices as the stress components. That is due to the staggered grid method, which prescribes different postions in the modeling domain for the displacements, stresses and elastic parameters, according to Virieux (1986).

The equations for the displacements in the x and z directions for the grid points near the free surface can be calculated in a similar fashion and will not be shown here.

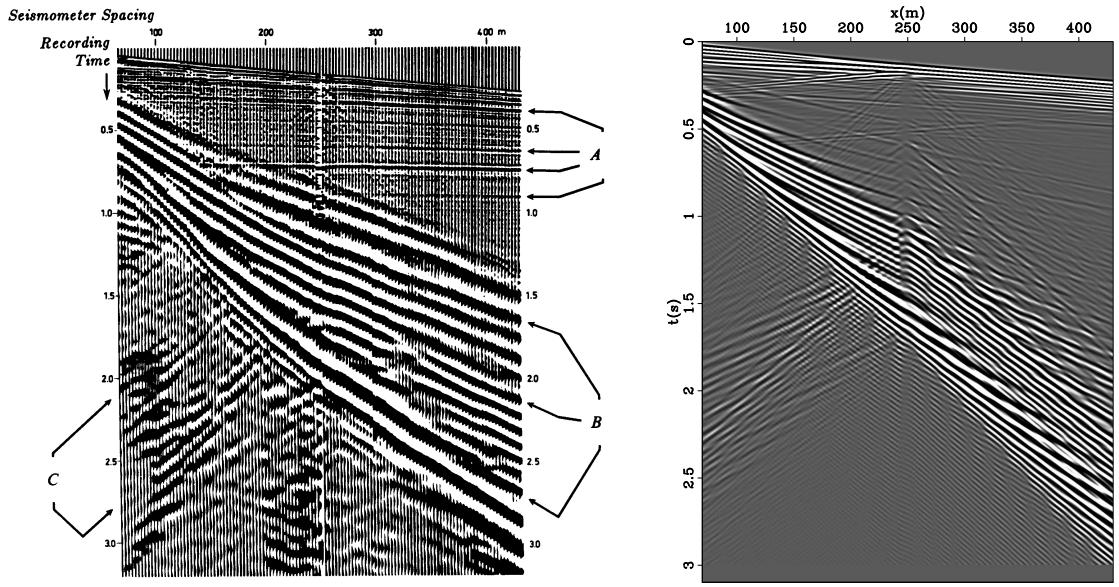
RESULTS

The method was used to create a 2D synthetic common shot gather that simulated a field shot gather. The example was taken from Claerbout (2011) and contains surface related events such as ground roll and backscattered waves from what appears to be a near-surface anomaly.

Figures 1(a) and 1(b) show the field data and synthetic data, respectively. The field data common-shot gather shows (A) hyperbolic events due to reflections within a thin shallow layer, (B) high amplitude dispersive surface waves (ground roll) and (C) backscattered

waves created by a near surface discontinuity. It is also possible to see forward scattered converted waves from the same discontinuity. In the synthetic example, the same events are observed and with the same kinematics. The hyperbolas at the top are much weaker in the synthetic example, due to a very strong contrast inside the low velocity layer. The dynamics of the example could be improved by manually adjusting the model properties.

The elastic parameters used to create the synthetic data emulate a free surface over a low velocity layer with steeply increasing shear velocities in the vertical direction and a near surface scatterer at $x = 250m$. Figures 2(a), 2(b) and 2(c) show the models for V_p, V_s and ρ, respectively. The modeling parameters are $\Delta x = \Delta z = 0.5m$ and $\Delta t = 3.27 \times 10^{-5}$. I used an explosive source with a peak frequency of 25 Hz. Although these grid parameters are finer than those usually employed in finite difference modeling at this frequency, they are required to avoid numerical dispersion of the slower surface modes.



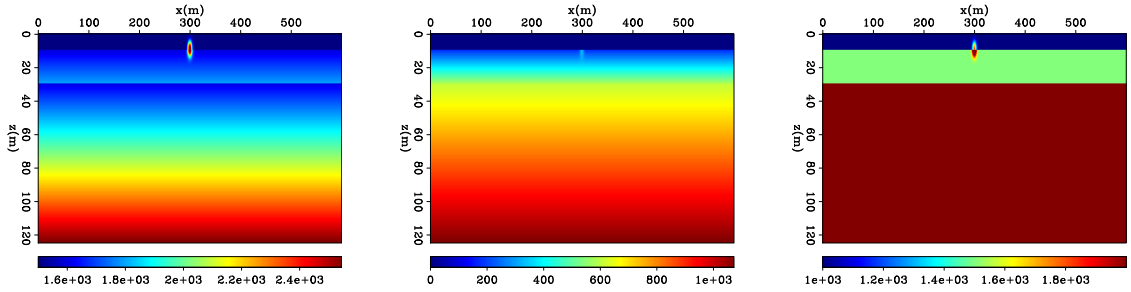
(a) (b)

Figure 1: (a) 2D field data and (b) synthetic data. The events marked as A, B and C correspond to primaries, ground roll and backscattered waves, respectively. Kinematically, there is a good match between the events observed in each gather, including the mode conversions from P to S waves at $x = 250m$. [**NR**] [**ER**] gcalves2/. shell,seisp-bei

DISCUSSION

A free-surface boundary condition was successfully implemented for a 10^{th} order in space and 2^{nd} order in time finite-difference stencil. The results show that the algorithm correctly recreates surface waves and that they are kinematically similar to those observed in a field data gather. The dynamics of the field data appear to have been reproduced, although a better correspondence might be achieved if an elastic model for the field data was available.

Due to the slower speeds of propagation for surface waves, the spatial sampling of the model was refined to avoid numerical dispersion. Consequently, the time sampling of the modeling must also be refined to maintain numerical stability.



(a) (b) (c)

Figure 2: (a)V_p, (b) V_s and (c) ρ models. [**ER**] gcalves2/. vp-bei,vs-bei,rho-bei

ACKNOWLEDGMENT

I would like to thank the Stanford Exploration Project sponsors for the ongoing support and Petrobras for support of my PhD.

REFERENCES

Alves, G. and B. Biondi, 2014, High-order elastic finite-difference modeling: Stanford Exploration Project Report, **152**, 327–340.

Boiero, D., E. Wiarda, and P. Vermeer, 2013, Surface-and guided-wave inversion for near-surface modeling in land and shallow marine seismic data: The Leading Edge, **32**, 638–646.

Boustani, B., S. Torabi, A. Javaherian, and S. A. Mortazavi, 2013, Ground roll attenuation using a curvelet-svd filter: a case study from the west of iran: Journal of Geophysics and Engineering, **10**, 055006.

Claerbout, J. F., 2011, Basic earth imaging.

Ikelle, L. and L. Amundsen, 2005, Introduction to petroleum seismology: Society of Exploration Geophysicists Tulsa, OK.

Liu, Y. and M. K. Sen, 2009, An implicit staggered-grid finite-difference method for seismic modelling: Geophysical Journal International, **179**, 459–474.

Takeuchi, H. and M. Saito, 1972, Seismic surface waves: Methods in computational physics, **11**, 217–295.

Virieux, J., 1986, P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method: Geophysics, **51**, 889–901.

Derived rotational and pressure-gradient seismic ocean-bottom data

Ohad Barak, Robert Brune, Paul Milligan and Shuki Ronen

ABSTRACT

In the Moere Vest acquisition a group of ocean-bottom nodes were deployed with a nominal spacing of two meters. We preprocessed the data of one shot line that traverses directly above these nodes. We then generated rotational and pressure-gradient data by differencing the geophone and hydrophone data of the adjacent nodes. We discuss the possibility of reverse-time propagation of such multicomponent seismic data.

INTRODUCTION

Rigid bodies in a three dimensional world have six degrees of freedom: three components of linear motion and three components of rotation. In the subsurface, the linear motions are the particle velocities and the rotations are the pitch, roll and yaw, as shown in the following table:

Axis	Displacement		Rotation	
Z	Vertical	v_z	Yaw	r_z
X	Radial	v_x	Roll	r_x
Y	Transverse	v_y	Pitch	r_y

where v_i are particle velocities along the axis, and r_i are rotation rates around the axis.

In ocean-bottom node acquisition, multicomponent geophones that are coupled to the seafloor record the vertical and the two horizontal components of particle velocity. Additionally, a hydrophone records the divergence of the wavefield $P = \kappa\,(\nabla \cdot \vec{u})$, where $\vec{u}$ are particle displacements and κ is the bulk modulus of the water to which the hydrophones are coupled. Rotations are a measurement of the curl of the wavefield $\vec{r} = \frac{1}{2}(\nabla \times \vec{v})$. However, as of yet there are no industry-grade solutions for recording rotational motion on the ocean bottom, though a few such recording stations have been deployed previously by Pillet et al. (2009).

The Moere Vest data includes a group of 26 ocean-bottom nodes, a "microspread," which have a unique geometry in that they are spaced at 2 m intervals. We estimated the three-component rotational motion by differencing adjacent geophones of these microspread nodes. Geophone differencing as a method of estimating the rotational motion signal has been shown previously in Barak et al. (2014a) and Edme et al. (2014). In the case of the microspread, the short 2 m interval between receivers ensures that most of the data is sampled well enough to prevent spatial aliasing, therefore, we assume that a differencing of the data recorded by adjacent nodes pertains to differences of displacements within half a wavelength.

Vassallo et al. (2012) use hydrophones together with pressure gradient sensors in marine streamer acquisition to interpolate the pressure wavefield in the crossline direction, between streamer cables. By differencing the hydrophone components of the adjacent nodes of the microspread, we similarly generated pressure gradient data. We can only obtain the two horizontal pressure gradients, since the nodes are all spread on the ocean-bottom, with no vertical separation.

Any differencing of data coming from physically separate sensors must be done with the caveat that we are in effect decreasing the signal to noise ratio in the resulting data. Each sensor may have a different coupling to the medium, reducing the reliability of the difference signal. Also, the data and the data difference are not collocated in space. Therefore, the data resulting from sensor differencing must be treated with a measure of suspicion. Barak et al. (2014b) discuss some of the problems associated with geophone differencing. Eventually, the only way to obtain a reliable recording of any physical quantity is to design a sensor that can measure that particular quantity at one point in space. The resulting rotational and pressure-gradient data can only be considered as an estimate of these data which would have been recorded with the adequate instrumentation. We would prefer not to use closely spaced receivers to measure rotations or pressure gradients in the field, but we do so here as a result of the special geometry of the microspread and the lack of the appropriate sensors.

Estimating rotational motion from geophone data

The stress-displacement relation for tangential stresses reads:

$$\sigma_{ij} = \mu \left(\partial_j u_i + \partial_i u_j \right), \tag{1}$$

where σ_{ij} are the tangential stresses, u_i are particle displacements and μ is the shear modulus.

At a free surface, or when going from a medium with shear strength to one without shear strength (such as the ocean-bottom interface), the tangential stresses σ_{ij} are zero. Therefore, assuming we have receivers laid out on a flat, horizontal sea bottom, we have

$$\begin{aligned} \partial_z u_y &= -\partial_y u_z, \\ \partial_z u_x &= -\partial_x u_z, \end{aligned} \tag{2}$$

meaning that the vertical derivative of the horizontal displacement component is equal to the horizontal derivative of the vertical displacement component.

Rotation is defined as the curl of the wavefield. Since our geophones record the time derivative of displacement (particle velocity), we use the time derivative of rotation, or rotation rate:

$$\vec{r} = \frac{1}{2}(\nabla \times \vec{v}) = \frac{1}{2}\left(\hat{X} \left(\partial_y v_z - \partial_z v_y \right) + \hat{Y} \left(\partial_z v_x - \partial_x v_z \right) + \hat{Z} \left(\partial_x v_y - \partial_y v_x \right) \right). \tag{3}$$

Substituting equation 2 into 3, we see that on the sea bottom

$$\begin{aligned} r_x &= \partial_y v_z, \\ r_y &= -\partial_x v_z, \\ r_z &= \frac{1}{2}\left(\partial_x v_y - \partial_y v_x\right), \end{aligned} \tag{4}$$

i.e., the horizontal rotation-rate components can be derived from the vertical geophones, and the vertical rotation-rate component can be derived from the horizontal geophones.

First-order in space elastic wave equation

The elastic wave equation for a homogeneous isotropic medium reads

$$(\lambda+\mu)\,\nabla\left(\nabla\cdot\vec{u}\right)+\mu\nabla^2\vec{u}=\rho\ddot{\vec{u}}, \tag{5}$$

where $\vec{u}$ are particle displacements, λ and μ are the Lamé parameters, and ρ is density. Using the vector identity $\nabla^2\vec{u}=\nabla\left(\nabla\cdot\vec{u}\right)-\nabla\times\left(\nabla\times\vec{u}\right)$, we get

$$(\lambda+2\mu)\,\nabla\left(\nabla\cdot\vec{u}\right)-\mu\nabla\times\left(\nabla\times\vec{u}\right)=\rho\ddot{\vec{u}}. \tag{6}$$

Since pressure $P=\kappa\nabla\cdot\vec{u}$, we may now write the elastic wave equation using only first derivatives in space:

$$\frac{\alpha^2}{\kappa}\nabla P-\beta^2\nabla\times\vec{r}=\ddot{\vec{u}}, \tag{7}$$

where α and β are P and S wave velocities respectively, P is pressure and $\vec{r}$ are the rotations. Using the notation $\vec{p}$ for the pressure gradient vector ∇P, we have

$$\frac{\alpha^2}{\kappa}\vec{p}-\beta^2\nabla\times\vec{r}=\ddot{\vec{u}}. \tag{8}$$

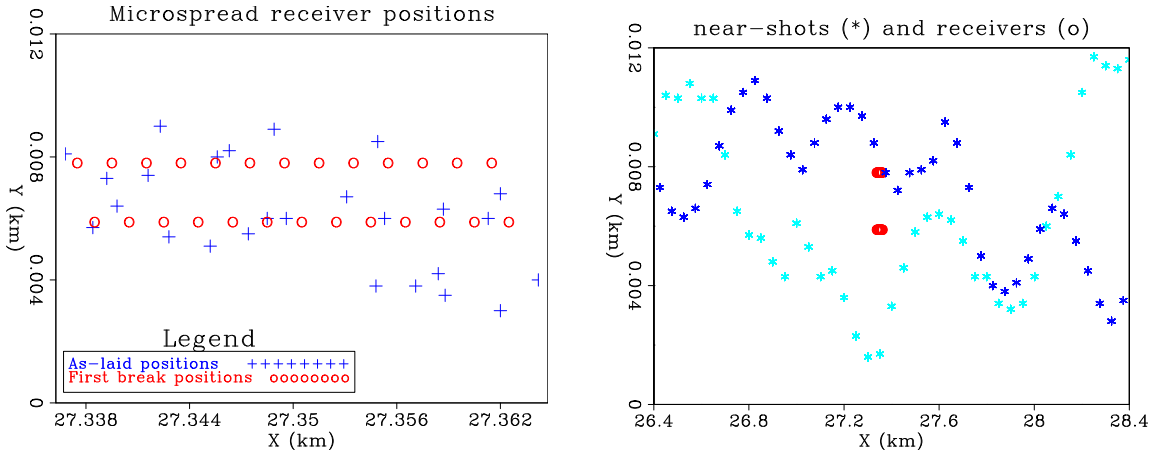
We have effectively modified the elastic wave equation to a system with first order derivatives in space, with three different measurable vector physical variables: pressure gradients, rotations and acceleration of displacements.

MICROSPREAD GEOMETRY

There were two sets of fields in the SEGY files that indicated receiver positions. One of them was the "as-laid" positions, which are the coordinates of the underwater Remotely Operated Vehicle (ROV) that deployed the nodes on the sea bottom. The other set of receiver positions were calculated using the first-break arrival time at each node from all shots in the survey. Figure 1(a) shows these two sets of receiver positions.

Despite the fact that the first-break positions appear to have been manually manipulated, we opted to use these node positions for the microspread for two reasons:

1. The ROV positioning system becomes less accurate as the depth of deployment increases. In this case, the water depth was 1.6km.
2. We have video footage of the deployment taken by the ROV, which shows that the ROV operator placed the nodes on a regular grid, with little deviations.



(a) (b)

Figure 1: (a) Microspread receiver "as-laid" positions vs. first-break positions. The length of the receiver line is 26 m, and the nominal receiver spacing is 2 m. (b) Microspread receiver positions (center of the figure, in red circles) with a subset of the shot positions from the two sail lines that are near the receivers. Note the exaggeration in the Y direction. The total shot line's length is 55 km. The largest crossline offset for these shot lines is 9 m, but over 90% of the shots have less than 5 m of crossline offset. Since the receivers are at a depth of 1.6 km, the acquisition is effectively 2D. ohad1/. gxgy-all,src-rec

Figure 1(b) shows the microspread array in relation to the shot positions of the near-offset shots. As can be seen from this figure there were two shot lines, passing almost directly above the nodes. The shot interval of each shot line is 50m, and they are interlaced so that the effective shot spacing is 25m. However, in the following data figures in this paper, we used only one of the shot lines, since tidal variations between the times the shot lines were acquired cause static shifts to appear on receiver gathers. The total shot line length was 55km, with the microspread nodes positioned more or less in the center of the line. The water depth was nearly 1.6km, and the crossline offset was at most 9m, therefore the survey is effectively 2D.

INSTRUMENT DESIGNATURE

Figure 2(a) shows the instrument response functions for the geophones and the hydrophone. In Figure 2(b) the spectra of the responses are shown. Note that the hydrophone's response is flat from around 4Hz, while that of the geophone is flat from around 8Hz. Consequently, we would expect the geophone data to exhibit more of a phase shift before and after the designature process.

Figure 3(a) shows a near-offset section of the hydrophone component of a receiver gather, after hyperbolic moveout was applied with water velocity. Blue wiggles indicate data before

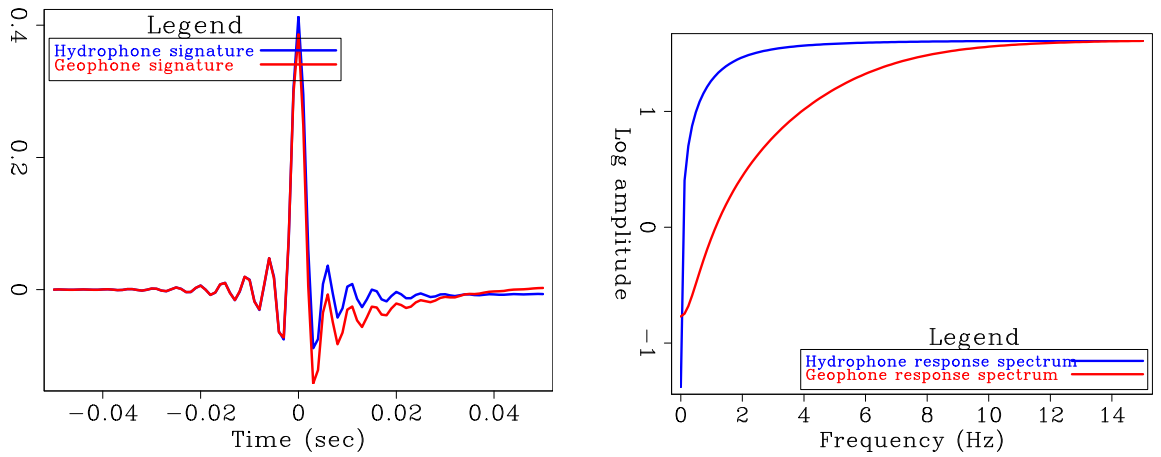


Figure 2: Instrument signatures and their spectra. (a) Hydrophone (blue) and geophone (red) instrument response function vs. time. (b) Hydrophone (blue) and geophone (red) instrument amplitude response vs. frequency. Note how the hydrophone's frequency response is flat starting from around 4Hz, while the geophone response flattens around 8Hz. Both instrument responses remain flat until 204Hz, where a high-cut filter was applied. [**ER**] ohad1/. CA-sig,CA-sigf

designature and red wiggles are after designature. Note how the designature caused the first break's energy to increase on the first negative lobe of the wavelet. There is also a slight phase shift concentrated on the first break. Figure 3(b) is vertical geophone gather before and after designature. The designature causes a much more marked change in this section. Much of the low frequency energy at $t = 1.1$s has been pushed up into the first break. Also, the low frequency bubble signal at $t = 1.19$s has shifted about 90°, and is now aligned with the bubble on the hydrophone section.

Figures 3(c) and 3(d) show the frequency spectrum of the stack of the moved out receiver gathers. The hydrophone's energy has been increased from about 4Hz and lower, while the vertical geophone's energy increased from 8Hz and lower. This is in accordance with the instrument frequency response in Figure 2(b).

A similar effect can be seen for the two horizontal geophones in Figures 4(c) and 4(d).

HORIZONTAL COMPONENT ALIGNMENT TO SURVEY COORDINATES

The horizontal components of the receivers of the microspread were not aligned with the survey's shot geometry. As each node is placed on the seafloor by the ROV, their alignment is not identical. For further processing, we required that the coordinates of the horizontal data components match those of the survey, and that they be consistent for all nodes. Consequently, a rotation of the horizontal components around the vertical axis was required. To gauge the amount of rotation, we first applied hyperbolic moveout to the data, then took a window of 100 milliseconds around the first break. We rotated the horizontal components

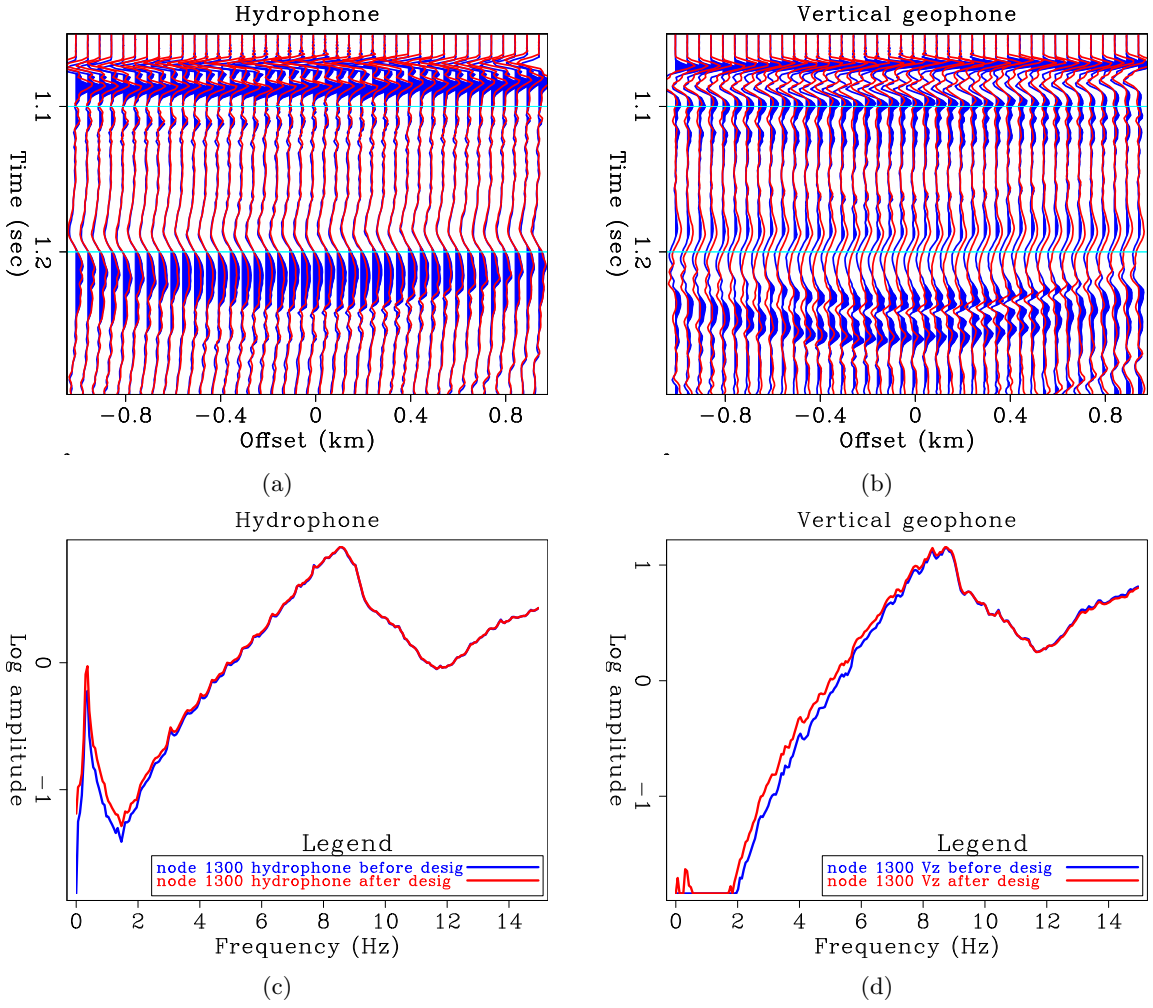


Figure 3: Data components of a receiver gather after hyperbolic moveout with water velocity, before and after instrument designature. (a) Hydrophone before (blue) and after (red) designature. (b) Vertical geophone before (blue) and after (red) designature. (c) and (d) are the log of the amplitude spectrum of the hydrophone and vertical geophone, respectively. Observe the increase in the lower amplitudes, particularly for the 8Hz geophone. The spike at 0.4Hz on the hydrophone is sea-swell ambient noise. Observe also how after designature, the low frequency bubble at $t = 1.19$s on the vertical geophone has an opposite polarity to the bubble on the hydrophone. Similarly, the direct arrival at $t = 1.08$s has an opposite polarity on the hydrophone vs. the vertical geophone. [**CR**]

ohad1/. comp-node1300-hyd,comp-node1300-vz,comp-node1300-hyd-f1,comp-node1300-vz-f1

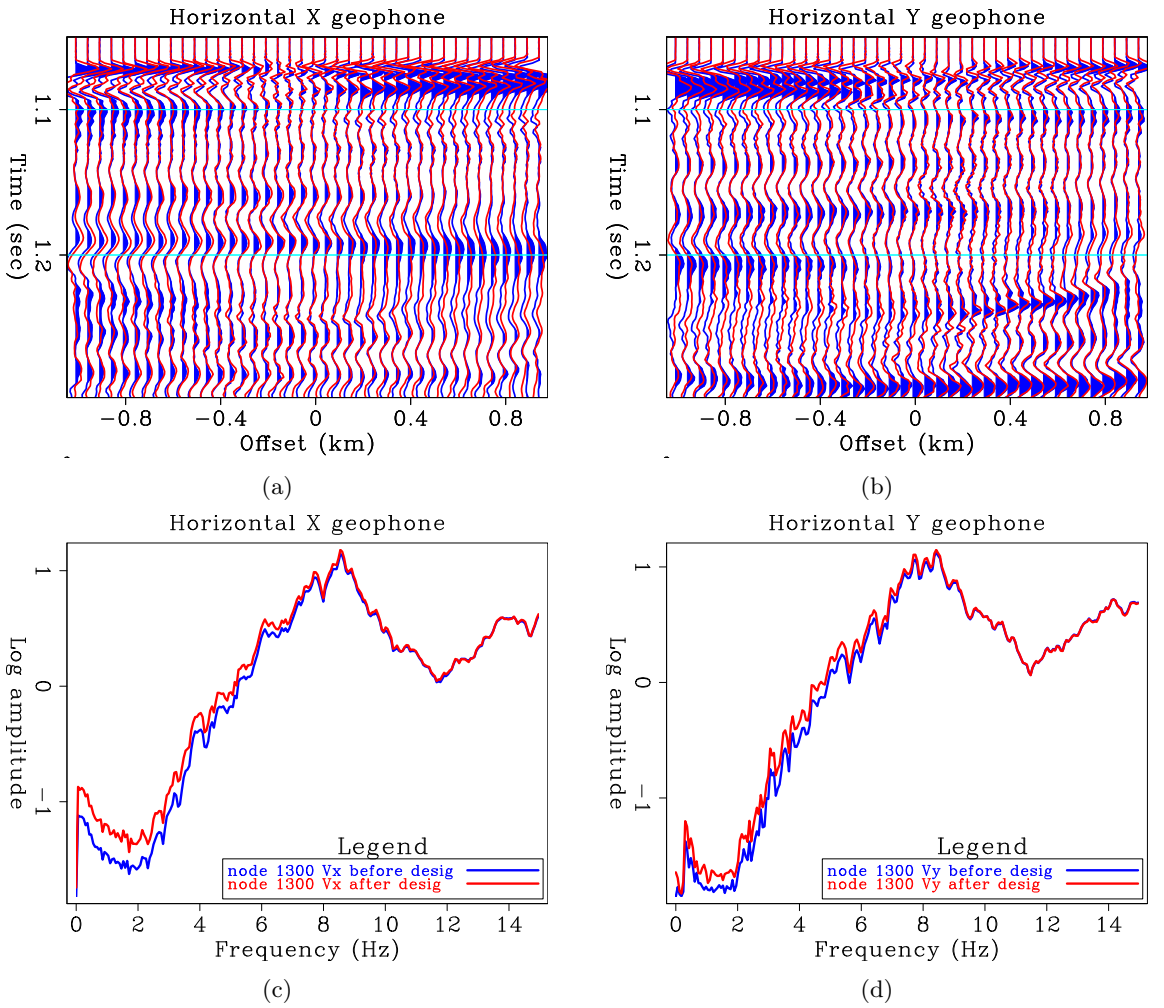
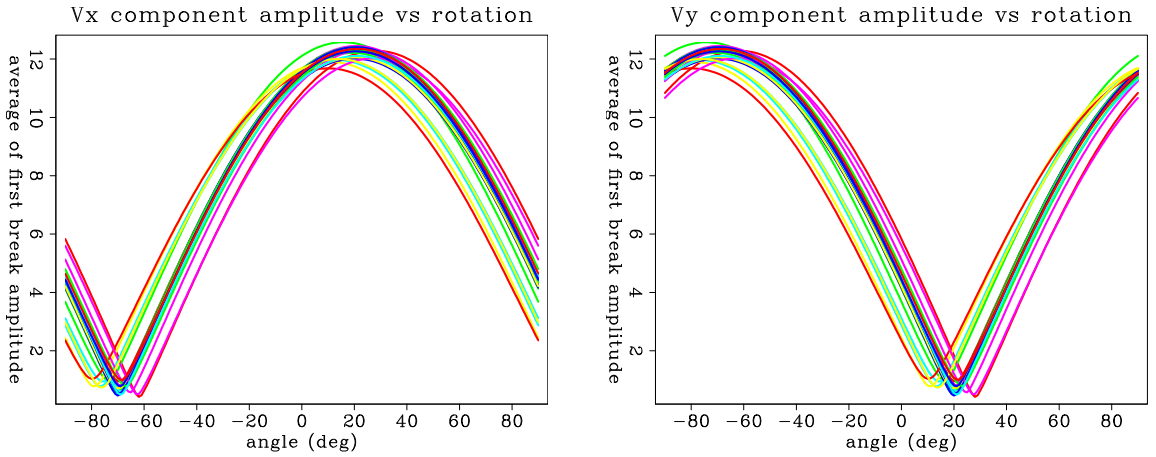


Figure 4: Data components of a receiver gather after hyperbolic moveout with water velocity, before and after instrument designature. (a) Horizontal inline geophone before (blue) and after (red) designature. (b) Horizontal crossline geophone before (blue) and after (red) designature. (c) and (d) are the log of the amplitude spectrum of these two component. Note that the horizontal geophones in this figure are not yet rotated to the true radial and transverse directions, however the geophone inline ('X') component is only about 20° away from the radial direction. [**CR**]

ohad1/. comp-node1300-vx,comp-node1300-vy,comp-node1300-vx-f1,comp-node1300-vy-f1

of each node, and then stacked the result and saw at what degree of rotation did we get a minimal amount of energy on the v_y component. The logic is that since the shot line has very little crossline offset with the microspread array, the first break should contain minimal energy in the crossline direction.

Figures 5(a) and 5(b) show the average amplitude of the first break on the two horizontal components of each node, as a function of rotation angle. We see that all the nodes appear to require a rotation of approximately 20^o to align them with the survey.



(a) (b)

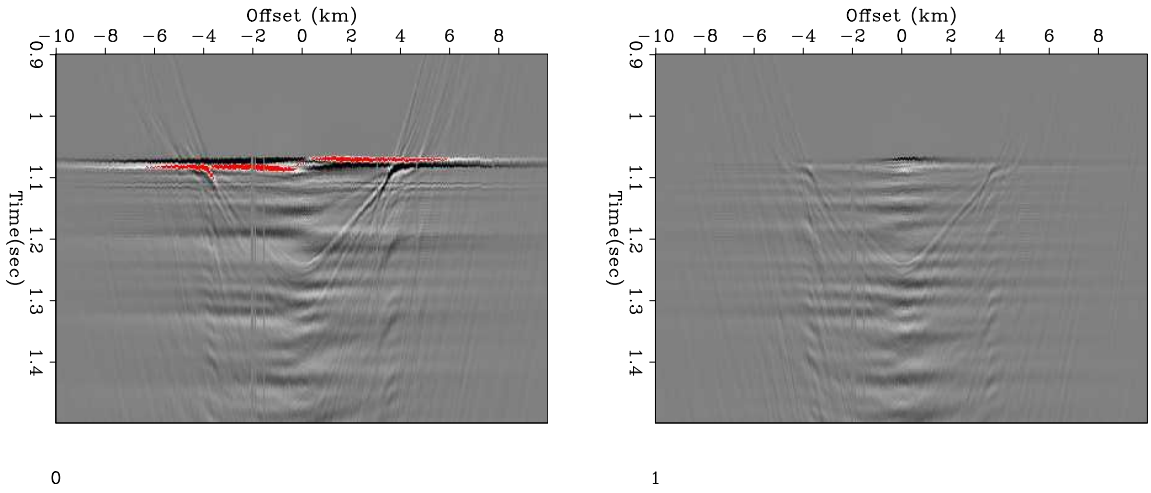
Figure 5: Average amplitude of a 100 millisecond window around the first break, as a function of the rotation of the horizontal components around the vertical axis. (a) v_x component. (b) v_y component. Each line in the graph is one of the 26 nodes of the microspread. It appears that in order to minimize the energy on the v_y component and thus have v_y perpendicular to the 2D shot line and v_x parallel to it, the nodes need to be rotated by about 20^o. [**CR**] ohad1/. hrot-allnode-sail1-vx,hrot-allnode-sail1-vy

Figure 6(a) is the v_y component before the rotation to survey coordinates, and Figure 6(b) is the same gather after rotation. Note how the first break's energy has been significantly weakened, along with much of the energy of the bubble.

GENERATION OF 9C DATA

To generate the three-component rotation-rate data, we used a finite-difference approximation to equations 4:

$$\begin{aligned}
r_y &\approx \frac{1}{\Delta x}\left(v_z^{i\Delta x,j\Delta y} - v_z^{(i+1)\Delta x,j\Delta y}\right), \\
r_x &\approx \frac{1}{\Delta y}\left(v_z^{i\Delta x,(j+1)\Delta y} - v_z^{i\Delta x,j\Delta y}\right), \\
r_z &\approx \frac{1}{4}\left[\frac{1}{\Delta x}\left(v_y^{i\Delta x,j\Delta y} - v_y^{(i+1)\Delta x,j\Delta y}\right) + \frac{1}{\Delta x}\left(v_y^{i\Delta x,(j+1)\Delta y} - v_y^{(i+1)\Delta x,(j+1)\Delta y}\right)\right] \\
&+ \frac{1}{4}\left[\frac{1}{\Delta y}\left(v_x^{i\Delta x,(j+1)\Delta y} - v_x^{i\Delta x,j\Delta y}\right) + \frac{1}{\Delta y}\left(v_x^{(i+1)\Delta x,(j+1)\Delta y} - v_x^{(i+1)\Delta x,j\Delta y}\right)\right]. \quad (9)
\end{aligned}$$



(a) (b)

Figure 6: The v_y component of the receiver gather at node 1300 after hyperbolic moveout with water velocity, (a) before alignment to crossline direction, and (b) after alignment to crossline direction. Since the survey geometry for these data was effectively 2D, the rotation should minimize the first break's ($t = 1.08$s) energy on the v_y component. [**CR**] ohad1/. rotcomp-node1300-vy1,rotcomp-node1300-vy2

Δx is the inline spacing between neighbouring nodes, and Δy is the crossline spacing. The order of subtraction operations in the differencing equations 9 were done so that the direction of the rotational axes would be the same as the linear geophone components, according to right-hand rule.

Figures 7(a),7(c) and 7(e) are the vertical (v_z), inline (v_x) and crossline (v_y) geophone components of the receiver gather of node 1730. Figures 7(b),7(d) and 7(f) are the yaw (r_z), roll (r_x) and pitch (r_y) rotational components. Notice that adjacent to each one of the geophone components is the rotational component around that geophone's axis.

We are not displaying the direct arrival and some of the associated bubbles that are between $t = 1.08$s and $t = 1.75$s. The water-bottom multiple appears at $t = 3.25$s, and can be seen well on the v_z and v_x sections.

The v_z component seems to contain mostly high-frequency reflections with a moveout consistent with P-wave velocity, but there are some lower frequency events apearing after every P reflection that have a much slower moveout. These events are commonly called "VZ noise", and may be caused by shear-wave scattering off the node body itself as a result of the incident P-wave. The v_x component contains mostly those shear-wave events, but some of the P-wave reflections apparent on v_z are also visible. The v_y component is much weaker than the other two geophone components, but a shear-wave event similar to the one on the v_x component at $t = 3.3$s is visible.

Observing the rotational components, we see that the one with the greatest energy is r_y. This fits with our expectation. Since the survey geometry is practically 2D, most of the linear motion should occur in the vertical and inline directions, which means that most of the rotational motion should occur around the crossline direction. Note also the generally

increased noise level on the rotational components.

Compare the v_z and r_y components, and observe how the P-waves are almost not visible on r_y, even though this section was obtained using two vertical geophones. This indicates that the P-waves generate a similar response on adjacent vertical geophones, and are removed by the differencing. Another way of saying the same thing is that the P-waves do not generate a rotational deformation of the surface. Instead, we see a section that is more similar to v_x, with events that have shear-wave moveouts (though slightly delayed compared to v_x). Shear waves, as their name suggests, generate a shear deformation of the surface, which expresses itself as rotational motion. Therefore, rotational data should preferentially record shear waves, and indeed the P events on the r_y component are much weaker than those visible on the v_z and v_x components.

The r_x rotational component seems to also contain some shear wave events related to the P-waves that hit the node. They are weaker than the events on r_y, indicating that if these are indeed the result of scattered shear waves, then these waves are causing mainly rotation around the crossline axis. The r_z section is the weakest of the rotations. This component should record events that cause a horizontal deformation around the vertical axis, i.e. Love or SH waves, neither of which seem particularly likely in this environment. What energy is on the r_z component seems to also be related to the incident P-waves.

We note that the nodes were not corrected for tilt. Each node is tilted approximately 2 to 4 degrees away from the vertical according to tiltmeters on the node bodies. Therefore, there may be leakage of the vertical motion into the horizontal geophones, and consequently into the r_z component.

To generate pressure-gradient data, we differenced the hydrophone component of the nodes:

$$\begin{aligned} p_x &\approx \frac{1}{\Delta x}\left(P^{(i+1)\Delta x, j\Delta y} - P^{i\Delta x, j\Delta y}\right), \\ p_y &\approx \frac{1}{\Delta y}\left(P^{i\Delta x, (j+1)\Delta y} - P^{i\Delta x, j\Delta y}\right). \end{aligned} \tag{10}$$

Figure 8(a) and 8(b) are the inline (p_x) and crossline (p_y) pressure gradients. Figure 8(c) is the hydrophone component. Note how the pressure gradients are much noisier than the hydrophone. This may be due to the P-waves having very low wavenumbers, so that there is not much of a difference between the hydrophone signal on adjacent nodes, and therefore differencing them increases the noise at the expense of the signal. This is especially true for the P_y component. If we consider the propagation path of the P-waves and the 2D survey geometry, we see that indeed there should not be a great difference in the pressure recorded by two nodes that are separated by 2 m in the crossline direction.

It is interesting that on the pressure-gradient sections we can see events with shear-wave moveouts at $t = 2.7$s that are not apparent on the hydrophone. This indicates that the shear events (whether they be caused by an actual shear reflection or a shear-wave scattering off the node body), are generating a pressure gradient without appearing on the pressure sensor as prominently as P-waves. This, in turn, indicates that some shear-wave energy is being recorded by the hydrophone.

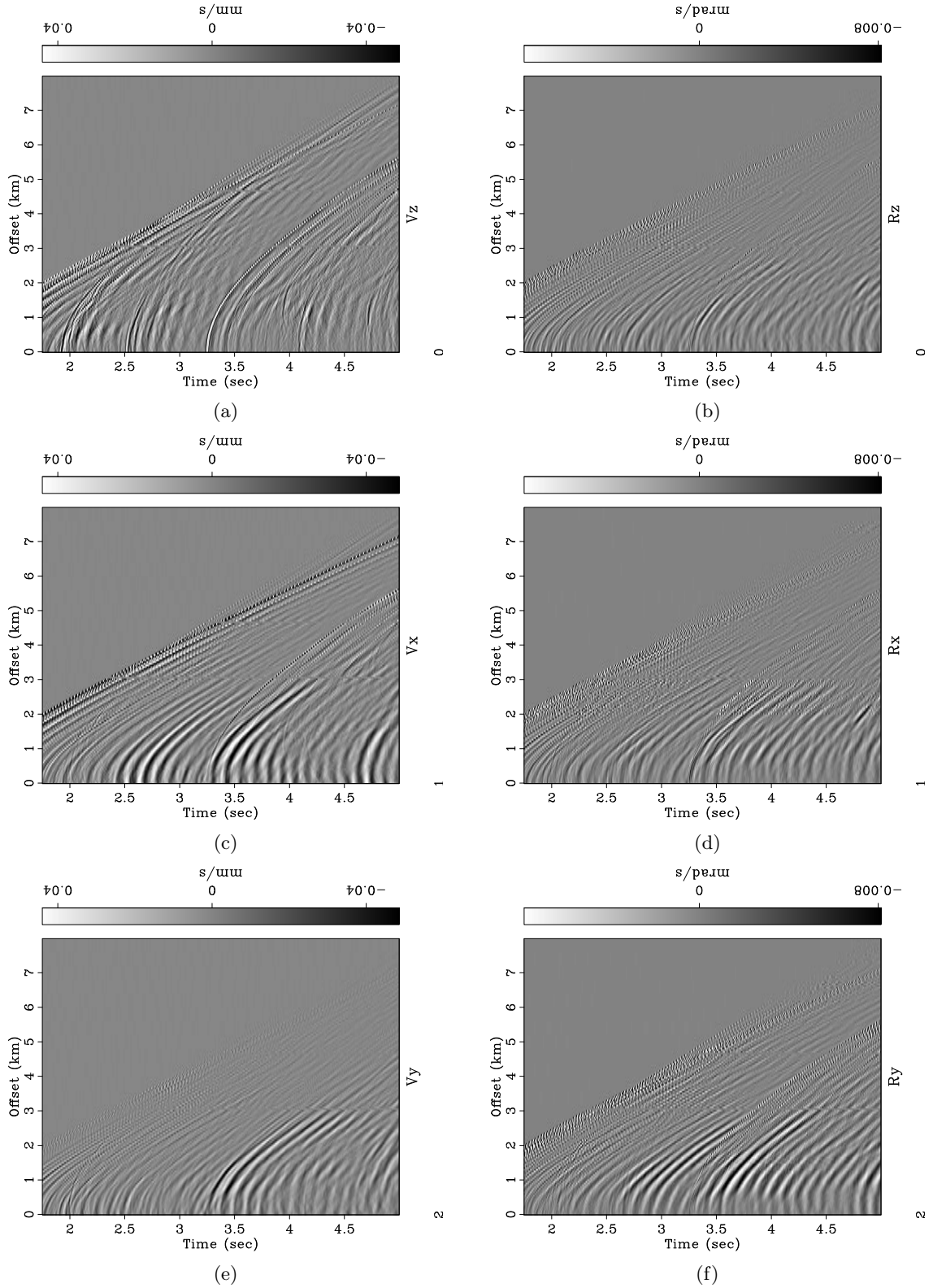


Figure 7: Six-component receiver gather of node 1730. (a) Vertical particle velocity v_z. (b) Yaw rotation r_z. (c) Inline particle velocity v_x. (d) Roll rotation r_x. (e) Crossline particle velocity v_y. (f) Pitch rotation r_y. [**CR**]

ohad1/. 6c-node1730-vz,6c-node1730-rz,6c-node1730-vx,6c-node1730-rx,6c-node1730-vy,6c-node1730-ry

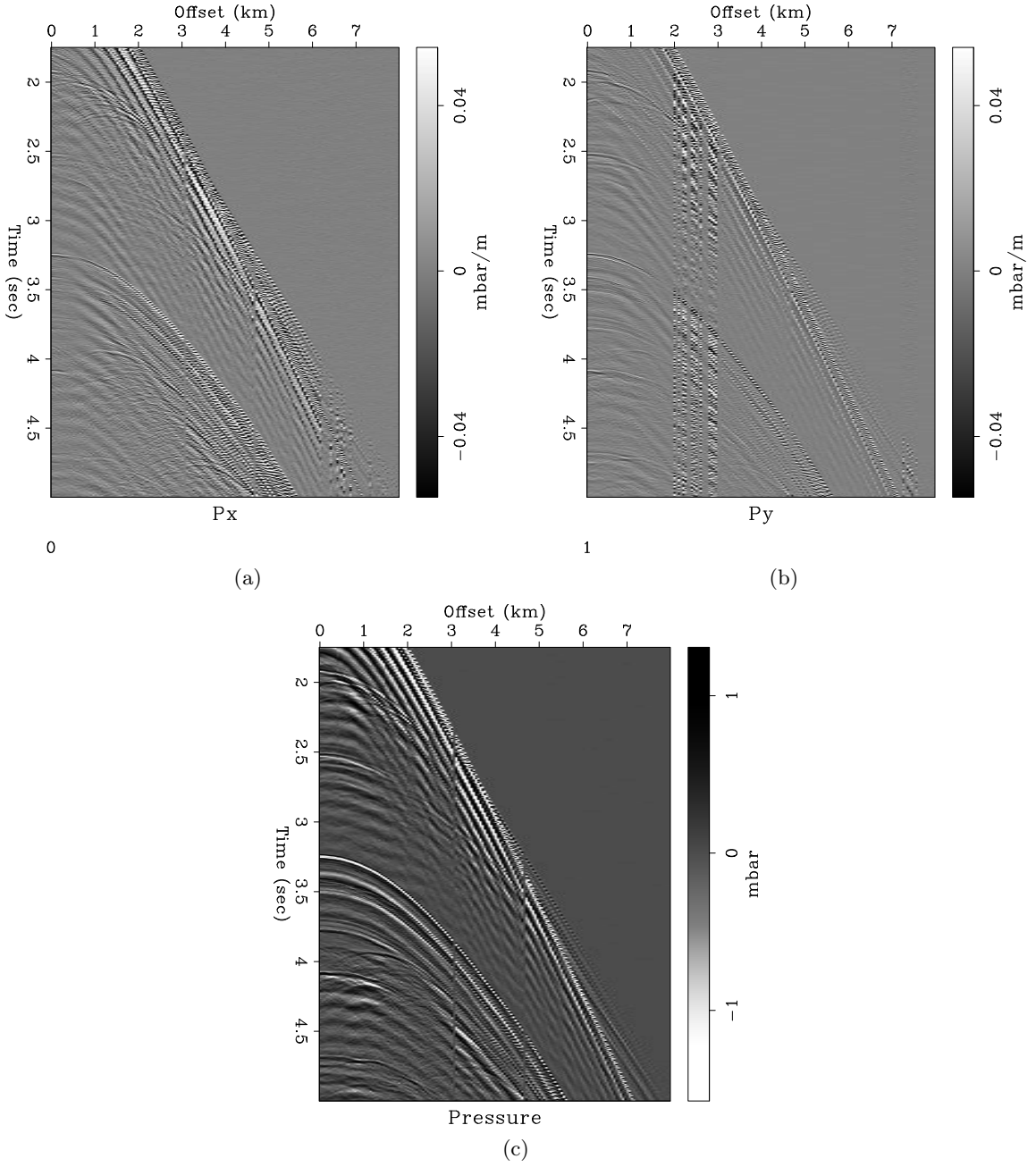


Figure 8: Pressure and pressure gradient for receiver gather of node 1730. (a) Pressure gradient in the inline direction P_x. (b) Pressure gradient in the crossline direction P_y. (c) Pressure. [**CR**] ohad1/. pg-node1730-px,pg-node1730-py,vol-node1730-hyd

DISCUSSION

In current seismic acquisition, geophones record only the displacements but not rotations, while hydrophones record the pressure but not its gradient. However, with the advent of a new generation of seismic sensors, these physical values will be measurable at each receiver position, giving us ten-component seismic data (1 pressure, 3 displacements, 3 rotations and 3 pressure gradients). Prototype pressure-gradient sensors are being developed, and rotation sensors exist and have seen some very limited use in seismic test surveys. Water accelerometers that effectively measure the pressure gradient in water have already been used widely in the exploration industry.

Consider an algorithm that propagates an elastic wavefield comprising the measured physical variables of displacements, pressure, pressure gradient and rotations (equation 8). Since both displacements and their spatial derivatives are recorded (pressure, pressure gradient and rotations), injecting them into an elastic wavefield propagator will generate waves that propagate not only in reverse time, but also in the opposite direction to their arrival at the receivers. This characteristic is used for deghosting acoustic marine streamer data by Vassallo et al. (2012). In the case of ocean-bottom seismic, this has the added advantage of separating the upgoing from the downgoing wavefield during elastic reverse-time propagation of the receiver data. This dismisses the need to separate the ocean-bottom data into upgoing and downgoing wavefields, using such methods as PZ summation (Barr and Sanders, 1989), which are commonly applied before imaging with ocean-bottom data (Wong et al., 2011).

Furthermore, injection of the displacement data and its derivatives will prevent mode conversion at the injection point, where P waves in the data are converted to S waves (and vice versa) immediately upon injection into the modeled wavefield. Such spurious modes can generate additional artifacts in the resulting image.

SUMMARY

We preprocessed seismic ocean-bottom node data so as to remove the receiver instrument signature and align the horizontal geophone components to the 2D survey coordinates. We used the fact that the receivers were deployed with small spacings to difference their data and estimate the rotational-motion and pressure-gradient data that would have been recorded had we instruments that were able to measure these physical variables directly on the ocean-bottom, thereby generating nine-component data. We anticipate that ten-component data comprising displacements, pressure, pressure gradient and rotations will improve seismic imaging with ocean-bottom data.

ACKNOWLEDGEMENTS

We thank Seabed Geosolutions for releasing the field data.

REFERENCES

Barak, O., S. de Ridder, J. Giles, P. Jaiswal, S. Ronen, and R. Brune, 2014a, Six-component seismic land data acquired with geophones and rotation sensors: Wave-mode separation using 6C SVD: SEG Technical Program Expanded Abstracts 2014, 1863–1867.

Barak, O., P. Jaiswal, S. de Ridder, J. Giles, R. Brune, and S. Ronen, 2014b, Six-component seismic land data acquired with geophones and rotation sensors: Wave-mode separation using 6C SVD : SEP-Report, **152**, 335–350.

Barr, F. and J. Sanders, 1989, Attenuation of water column multiples using pressure and velocity detectors in a water bottom cable: SEG Technical Program Expanded Abstracts 1989, 653–656.

Edme, P., E. Muyzert, and E. Kragh, 2014, Efficient land seismic acquisition sampling using rotational data: 76th EAGE Conference and Exhibition, Seismic Noise Attenuation Session.

Pillet, R., A. Deschamps, D. Legrand, J. Virieux, N. Bethoux, and B. Yates, 2009, Interpretation of broadband ocean-bottom seismometer horizontal data seismic background noise: Bulletin of the Seismological Society of America, **99**, 1333–1342.

Vassallo, M., K. Eggenberger, D. J. van Manen, K. Ozdemir, J. Robertsson, and A. Ozbek, 2012, Contributions of the horizontal and vertical components of particle velocity in 3D pressure wavefield reconstruction on dense receiver grids using generalized matching pursuit: SEG Technical Program Expanded Abstracts, 1–5.

Wong, M., S. Ronen, and B. Biondi, 2011, Leastsquares reverse time migration/inversion for ocean bottom data: A case study: SEG Technical Program Expanded Abstracts 2011, 2369–2373.

Pseudo-acoustic vertical transverse isotropic migration velocity analysis using two-way wavefield propagation

Carlo Fortini

ABSTRACT

Wave-Equation Migration Velocity Analysis (WEMVA) is widely used as a tool to reconstruct a model of the subsurface, such that some features of the migrated image are met. I show an anisotropic WEMVA workflow based on the Vertical Transverse Isotropic (VTI) approximation for the velocity model and a pseudo-acoustic anisotropic two-way wave-equation modeling engine. I derive the theory of WEMVA starting from the gradients of the anisotropic Full Waveform Inversion (FWI) that provides the input images for the velocity analysis. In doing so, I introduce the concept of *generalized images* that defines the FWI gradients computed with respect to the anisotropic parameters as different images of the subsurface. The results of some preliminary tests on the use of the generalized images as input for WEMVA suggest that this approach could help improving the accuracy and rate of convergence of WEMVA.

INTRODUCTION

Velocity model building is a key element in the context of seismic processing and is still one of the most challenging problems in the exploration industry. Currently, velocity building is conducted with techniques that work either in the *data-space* or *image-space*. To the first class belong all those algorithms that go under the name of Full Waveform Inversion (FWI); whereas, to the second class belong the so-called Wave Equation Migration Velocity Analysis (WEMVA) processes. There are various advantages that drive to the use of image-space techniques as opposed to data-space techniques: First, the migrated images are usually cleaner than recorded data. Moreover, the requirements for the initial model are less strict for WEMVA rather than FWI techniques.

Most of the time, WEMVA is conducted under an isotropic approximation of the subsurface model, neglecting its anisotropic characteristics. However, the increasing offset and azimuth in recent seismic data acquisition has heightened the need for an anisotropic parametrization of the velocity model. Neglecting the anisotropy can, in fact, lead to a wrong interpretation of the subsurface structures. In many cases, a Vertical Transverse Isotropic (VTI) approximation can be used to more accurately describe the subsurface. Li and Biondi (2011) showed that WEMVA can be successfully used to retrieve an anisotropic model using a one-way VTI wave-equation as propagation engine. Many authors (Duveneck et al. (2008); Fletcher et al. (2009); Zhang and Zhang (2009)) proposed migration and modeling algorithms schemes for VTI media, primarily based on the *pseudo-acoustic* approximation first proposed by Alkhalifah (1998). Because the anisotropic parameters are sensitive to events that propagate with large angles, the use of two-way wave-equation modeling algorithm can provide significant improvement to the results of anisotropic WEMVA

(Li et al., 2012).

In this paper, I show an anisotropic WEMVA workflow based on two-way wave-equation propagation engine. I first introduce briefly the wave-equation I use; and then, I derive the computation of the WEMVA gradients. In doing so, I start from the gradients of FWI that provide the input image for WEMVA. I also show some preliminary results on the use of the anisotropic generalized images as input for the image-space velocity model inversion.

ACOUSTIC VERTICAL TRANSVERSE ISOTROPIC WAVE-FIELD MODELING

To avoid complications coming from the presence of *shear waves* and for computational efficiency, wave propagation modeling with two-way wave-equation is usually conducted under the *pseudo-acoustic* VTI approximation (Alkhalifah, 1998). This approximation consists of setting the shear-wave velocity to zero in the exact elastic wave equations and leads to the following system of partial differential equations (Duveneck et al., 2008):

$$\begin{cases} \dfrac{1}{v_p^2}\dfrac{\partial^2 p_h}{\partial t^2} = (1+2\varepsilon)\dfrac{\partial^2 p_h}{\partial x^2} + \sqrt{1+2\delta}\dfrac{\partial^2 p_v}{\partial z^2} + f_h \\ \dfrac{1}{v_p^2}\dfrac{\partial^2 p_v}{\partial t^2} = \sqrt{1+2\delta}\dfrac{\partial^2 p_h}{\partial x^2} + \dfrac{\partial^2 p_v}{\partial z^2} + f_v, \end{cases} \tag{1}$$

where $\epsilon(x,y,z)$, $\delta(x,y,z)$ and $v_p(x,y,z)$ are, respectively, the anisotropic Thomsen parameters (Thomsen, 1986) and the vertical P-wave velocity. $p_h(x,y,z,t)$ and $p_v(x,y,z,t)$ are the horizontal and vertical normal stresses. Similarly, $f_h(x,y,z,t)$ and $f_v(x,y,z,t)$ are the horizontal and vertical source terms. The system 1 can be re-written in a matrix-vector notation as follows:

$$\mathbf{L}(v_p,\varepsilon,\delta)\mathbf{p} = \mathbf{f}, \tag{2}$$

where $\mathbf{p} = [p_h, p_v]^T$, $\mathbf{f} = [f_h, f_v]^T$ is the source term and

$$\mathbf{L}(v_p,\varepsilon,\delta) = \begin{bmatrix} \frac{1}{v_p^2}\frac{\partial^2}{\partial t^2} - (1+2\varepsilon)\frac{\partial^2}{\partial x^2} & -\sqrt{1+2\delta}\frac{\partial^2}{\partial z^2} \\ -\sqrt{1+2\delta}\frac{\partial^2}{\partial x^2} & \frac{1}{v_p^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial z^2} \end{bmatrix}. \tag{3}$$

When both ϵ and δ are set to zero, system 1 is equivalent to the isotropic acoustic second-order wave-equation.

Because the equations previously presented have been derived with the acoustic VTI approximation, they are kinematically equivalent to acoustic VTI equations previously described in the literature (e.g., Fletcher et al. (2009); Zhang and Zhang (2009)) and share both the benefits and drawbacks of all the equations based on an acoustic approximation. They have the well-known problem of source-generated shear waves (Grechka et al., 2004), which for the purposes of P-wave modeling are regarded as artifacts. In my work, I always consider the case of an acquisition surface placed in an isotropic layer; and thus, the source-generated shear waves does not constitute a problem. Another consequence of the acoustic VTI approximation is the condition $\epsilon \geq \delta$ to ensure stability.

The system of equation in 1 is a special case of the one presented in Li et al. (2012) for the specific case of a constant density medium. I used these equations as the basis for an anisotropic two-way modeling and reverse-time migration implementation for all the examples in this paper.

IMAGING PRINCIPLE FOR VERTICAL TRANSVERSE ISOTROPIC MIGRATION

What is usually referred to as the subsurface image in the context of seismic migration can be computed as the first gradient of a Full Waveform Inversion (FWI) inverse problem (Tarantola, 1984), when the initial model does not contain any sharp contrasts. The objective function that is usually minimized for solving the FWI problem is defined in the *data-space* as follows:

$$J_{\text{FWI}} = \frac{1}{2} \left\langle d_{\text{obs}} - d_{\text{est}}, d_{\text{obs}} - d_{\text{est}} \right\rangle , \tag{4}$$

where d_{obs} is the recorded data and d_{est} is the data estimated using the wave-equation in 1 and the current models (ϵ, δ, v_p). d_{est} is simply obtained by the sampling of the wavefield $\mathbf{p}$ at the location of the original acquisition surface where d was recorded. At the very first iteration, we usually have only a smooth estimate for the subsurface models. If we only account for the reflection events (neglecting the diving waves, the refractions and the direct arrivals), then $d_{\text{est}} = 0$ because no reflection events can be generated when using a smooth model. The FWI gradient would try to introduce exactly those interfaces needed to explain the reflection events recorded and observed in the data giving as output what we usually call "migrated image."

Li et al. (2012) show that the gradient of the FWI objective function can be computed using the *adjoint-state method*. The final result is the following:

$$\nabla_{\mathbf{m}} J_{\text{FWI}} = - \left\langle \mathbf{q}, \frac{\partial \mathbf{L}}{\partial \mathbf{m}} \mathbf{p} \right\rangle , \tag{5}$$

where $\langle , \rangle$ indicates the scalar product in time and $\mathbf{q} = [q_h(x, y, z, t), q_v(x, y, z, t)]$ is the wavefield computed by back-propagating the data residual $d_{\text{obs}} - d_{\text{est}}$. $\mathbf{m}$ indicates the vector of the model parameters that describe the subsurface. Depending on the case under analysis, $\mathbf{m}$ can be either 1D or multidimensional. For the isotropic case, $\mathbf{m} = [v_p]$; while for the anisotropic scenario described in the previous section, $\mathbf{m}$ is the vector of the three components of the anisotropic subsurface model ($\mathbf{m} = [v_p, \epsilon, \delta]^T$). Equation 5 can be expanded as

$$\begin{aligned} \nabla_{v_p} J_{\text{FWI}} &= - \left\langle \mathbf{q}, \frac{\partial \mathbf{L}}{\partial v_p} \mathbf{p} \right\rangle ; \\ \nabla_{\varepsilon} J_{\text{FWI}} &= - \left\langle \mathbf{q}, \frac{\partial \mathbf{L}}{\partial \varepsilon} \mathbf{p} \right\rangle ; \\ \nabla_{\delta} J_{\text{FWI}} &= - \left\langle \mathbf{q}, \frac{\partial \mathbf{L}}{\partial \delta} \mathbf{p} \right\rangle , \end{aligned} \tag{6}$$

where

$$\frac{\partial \mathbf{L}}{\partial v_p} = \begin{bmatrix} -\frac{2}{v_p^3}\frac{\partial^2}{\partial t^2} & 0 \\ 0 & -\frac{2}{v_p^3}\frac{\partial^2}{\partial t^2} \end{bmatrix}; \tag{7}$$

$$\frac{\partial \mathbf{L}}{\partial \epsilon} = \begin{bmatrix} -2\frac{\partial^2}{\partial x^2} & 0 \\ 0 & 0 \end{bmatrix}; \tag{8}$$

$$\frac{\partial \mathbf{L}}{\partial \delta} = \begin{bmatrix} 0 & -\frac{1}{\sqrt{1+2\delta}}\frac{\partial^2}{\partial z^2} \\ -\frac{1}{\sqrt{1+2\delta}}\frac{\partial^2}{\partial x^2} & 0 \end{bmatrix}. \tag{9}$$

The explicit expressions for the computation of the three gradients are

$$\begin{aligned} \nabla_{v_p} J_{\text{FWI}} = I_{v_p} &= \frac{2}{v_p^3}\left(\int \frac{\partial^2 p_h}{\partial t^2} q_h dt + \int \frac{\partial^2 p_v}{\partial t^2} q_v dt\right); \\ \nabla_{\varepsilon} J_{\text{FWI}} = I_{\epsilon} &= 2\int \frac{\partial^2 p_h}{\partial x^2} q_h dt; \\ \nabla_{\delta} J_{\text{FWI}} = I_{\delta} &= \frac{1}{\sqrt{1+2\delta}}\left(\int \frac{\partial^2 p_h}{\partial x^2} q_h dt + \int \frac{\partial^2 p_v}{\partial z^2} q_v dt\right). \end{aligned} \tag{10}$$

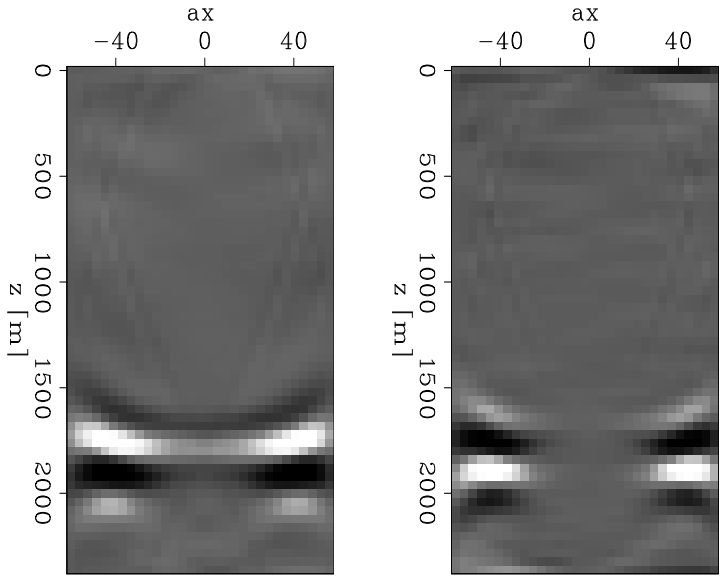
In analogy with the isotropic acoustic case in which the first FWI gradient is referred as "the image", I refer to the three components of the gradient of the VTI FWI objective function (I_{v_p}, I_ϵ and I_δ) as *generalized images*. In fact, the expressions in equation 10 can be seen as generalized imaging principles for the VTI migration. From a kinematic point of view, they all contain the same events: They all give an image of the same subsurface interfaces. However, they are not equivalent in term of amplitudes and illumination of the subsurface reflectors. Figure 1 shows two angle gathers extracted from, respectively, I_{v_p} and I_ϵ for a simple synthetic example. The models used to both generate and compute the FWI gradients are all constant (v_p, ϵ and δ) with a single, sharp discontinuity that simulates a flat reflector positioned at $z = 1,800$ m.

The main difference between the two images is the lack of illumination at zero-incidence angle for the case of I_ϵ. This behavior is in accordance with the theory: Events traveling almost vertically have no sensitivity to the anisotropy parameter ϵ.

ANISOTROPIC WAVE EQUATION MIGRATION VELOCITY ANALYSIS

Wave Equation Migration Velocity Analysis (WEMVA) is a nonlinear inversion process that aims at estimating a background (anisotropic) velocity model, such that a pre-selected objective function is minimized. Unlike FWI, the objective function for WEMVA is defined in the *image-space*. The generic objective function for WEMVA is

$$J_{\text{WEMVA}} = \frac{1}{2}\left\langle PI, PI \right\rangle, \tag{11}$$



(a) (b)

Figure 1: Angle gathers extracted from (a) I_{v_p} and (b) I_ϵ. Notice the differences in the illumination. [**NR**] carlo1/. test1-angveltrue,test1-angepstrue

where I is a seismic image and P is a *penalty* operator. P is usually chosen, such that the minimization of the objective function leads to a seismic image with some desired features. The most commonly used objective functions are

$$\begin{aligned} J_{\text{DSO}} &= \frac{1}{2}\left\langle \mathbf{h}I(\mathbf{x},\mathbf{h}), \mathbf{h}I(\mathbf{x},\mathbf{h})\right\rangle; \\ J_{\text{PSM}} &= \frac{1}{2}\left\langle I(\mathbf{x},\mathbf{h}=0), I(\mathbf{x},\mathbf{h}=0)\right\rangle. \end{aligned} \tag{12}$$

I is expressed as a function of the position $\mathbf{x} = (x, y, z)$ and the subsurface offset shift (lag) $\mathbf{h} = [h_x, h_y]$ (Sava and Fomel, 2006). The former is the so-called *Differential Semblance Optimization* (DSO), while the latter is the *Power of Stack Migration* (PSM) (Symes and Kern (1994); Toldi (1989); Shen (2004)). The images used as input for WEMVA can be computed as shown in the previous section, solving the first step of the FWI. In particular, the image as a function of the Subsurface Offset Domain Common Image Gathers (SOCIG) can be computed as (for the case of I_{v_p})

$$I_{v_p}(\mathbf{x},\mathbf{h}) = \frac{2}{v_p^3}\left(\int \frac{\partial^2(S_{+\mathbf{h}}p_h)}{\partial t^2}(S_{-\mathbf{h}}q_h)dt + \int \frac{\partial^2(S_{+\mathbf{h}}p_v)}{\partial t^2}(S_{-\mathbf{h}}q_v)dt\right), \tag{13}$$

where $S_{+\mathbf{h}}$ is a shifting operator that shifts the wavefield by $+\mathbf{h}$ in the $\mathbf{x}$ direction. Similarly, the operator $S_{-\mathbf{h}}$ shifts the wavefield in the opposite direction. Note that $(S_{+\mathbf{h}})^* = S_{-\mathbf{h}}$. $I_\epsilon(\mathbf{x},\mathbf{h})$ and $I_\delta(\mathbf{x},\mathbf{h})$ can be computed similarly. When the correct model is used for the construction of the seismic images, all the energy is focused at zero-lag ($\mathbf{h} = 0$). Moreover, the image extracted at $\mathbf{h} = 0$ corresponds to the one that can be obtained by stacking

(along the angles dimension) the seismic image decomposed in angle gathers. The DSO objective function measures the nonfocused energy in the subsurface-offset-gathers and is thus minimized when performing WEMVA. On the other hand, the PSM measures the energy of the stack along the angles of the seismic image and has thus to be maximized.

I now show the computation of the gradients of the WEMVA inverse problem when the objective function is the DSO and using I_{v_p} as input. The same derivation can be used also to compute the gradients for the case of the PSM objective function. I follow a symbolic derivation based on the *adjoint-state* method (Plessix, 2006) that can be found in Li et al. (2012). For the sake of the notation, I indicate with $I_{\mathbf{h}}$ the generalized image $I_{v_p}(\mathbf{x}, \mathbf{h})$.

The first step is the definition of the Lagrangian augmented functional as

$$\begin{aligned}\mathcal{L}(\mathbf{p}, \mathbf{q}, I_{\mathbf{h}}, \lambda, \mu, \gamma_{\mathbf{h}}, \mathbf{m}) = &\sum_{\mathbf{h}} \frac{1}{2} \langle \mathbf{h} I_{\mathbf{h}}, \mathbf{h} I_{\mathbf{h}} \rangle \\ &+ \langle \lambda, \mathbf{f} - \mathbf{L}\mathbf{p} \rangle \\ &+ \left\langle \mu, \mathbf{f}' - \mathbf{L}^* \mathbf{q} \right\rangle \\ &+ \sum_{\mathbf{h}} \left\langle \gamma(\mathbf{x}, \mathbf{h}), \frac{\partial \mathbf{L}^*}{\partial v_p} (S_{+\mathbf{h}} \mathbf{p})^* (S_{-\mathbf{h}} \mathbf{q}) - I_{\mathbf{h}} \right\rangle.\end{aligned} \tag{14}$$

The adjoint-state equations can then be computed as

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathbf{p}} &= -\mathbf{L}^* \lambda + \sum_{\mathbf{h}} \frac{\partial \mathbf{L}}{\partial v_p} (S_{+\mathbf{h}})^* (S_{-\mathbf{h}} \mathbf{q}) \gamma_{\mathbf{h}} = 0; \\ \frac{\partial \mathcal{L}}{\partial \mathbf{q}} &= -\mathbf{L}\mu + \sum_{\mathbf{h}} \frac{\partial \mathbf{L}}{\partial v_p} (S_{-\mathbf{h}})^* (S_{+\mathbf{h}} \mathbf{p}) \gamma_{\mathbf{h}} = 0; \\ \frac{\partial \mathcal{L}}{\partial I_{\mathbf{h}}} &= -\gamma_{\mathbf{h}} + \mathbf{h}^2 I_{\mathbf{h}} = 0.\end{aligned} \tag{15}$$

The solution of the adjoint-state equations allows the retrieval of the the adjoint-state variables $(\lambda(\mathbf{x}, t), \mu(\mathbf{x}, \mathbf{t}), \gamma(\mathbf{x}, \mathbf{h}))$. The gradients of the objective function in equation 12 with respect to the three anisotropy parameters then can be computed as follows:

$$\begin{aligned}\nabla_{v_p} J_{\text{WEMVA}} &= \left\langle \lambda, -\frac{\partial \mathbf{L}}{\partial v_p} \mathbf{p} \right\rangle + \left\langle \mu, -\frac{\partial \mathbf{L}^*}{\partial v_p} \mathbf{q} \right\rangle; \\ \nabla_{\epsilon} J_{\text{WEMVA}} &= \left\langle \lambda, -\frac{\partial \mathbf{L}}{\partial \epsilon} \mathbf{p} \right\rangle + \left\langle \mu, -\frac{\partial \mathbf{L}^*}{\partial \epsilon} \mathbf{q} \right\rangle; \\ \nabla_{\delta} J_{\text{WEMVA}} &= \left\langle \lambda, -\frac{\partial \mathbf{L}}{\partial \delta} \mathbf{p} \right\rangle + \left\langle \mu, -\frac{\partial \mathbf{L}^*}{\partial \delta} \mathbf{q} \right\rangle.\end{aligned} \tag{16}$$

Generalized images

As previously discussed, I consider herein the model for the subsurface to be characterized by the three Thomsen parameters v_p, ϵ, δ. Equation 10 shows that in this case, the seismic images at our disposal are three as well. Regardless the choice of the specific cost function (either DSO or PSM), it is possible to minimize/maximize the objective function computed with any of the three images by computing the gradients with respect to the three parameters. In the derivation followed in the previous section, for instance, I derived the gradients of WEMVA for v_p, ϵ and δ by using I_{v_p} as input. This input would correspond to find the anisotropic model, such that the image I_{v_p} has the desired characteristics. As a matter of fact, anisotropic WEMVA is often performed using only I_{v_p} and inverting for all the anisotropic parameters (Weibull and Arntsen (2014); Li et al. (2012)).

However, it is also possible to use the correspondent image when trying to invert for a specific parameter. This would mean, for instance, to minimize/maximize the DSO/PSM computed using I_{v_p} when inverting for v_p, I_ϵ for ϵ and I_δ for δ. The choice of the image to be used as input affects the computation of the WEMVA gradient. As a matter of fact, the seismic image chosen as input for WEMVA plays a role in the computation of the adjoint-state variables (equation 16). It is reasonable to expect the associated image to have a greater sensitivity to the correspondent image rather than the image computed for another parameter. All the parameters affect all the three images but each parameter produces a first-order effect to the correspondent image, and a second-order effect to the other two. If, for instance, I_{v_p} is used as input to invert for all the parameters, the inaccuracies in the ϵ and δ model would produce second-order effects on the image and they would thus be less significant than the effects caused by inaccuracies in the velocity model.

In the next section, I show the result of two preliminary tests I performed to validate the previous assumptions.

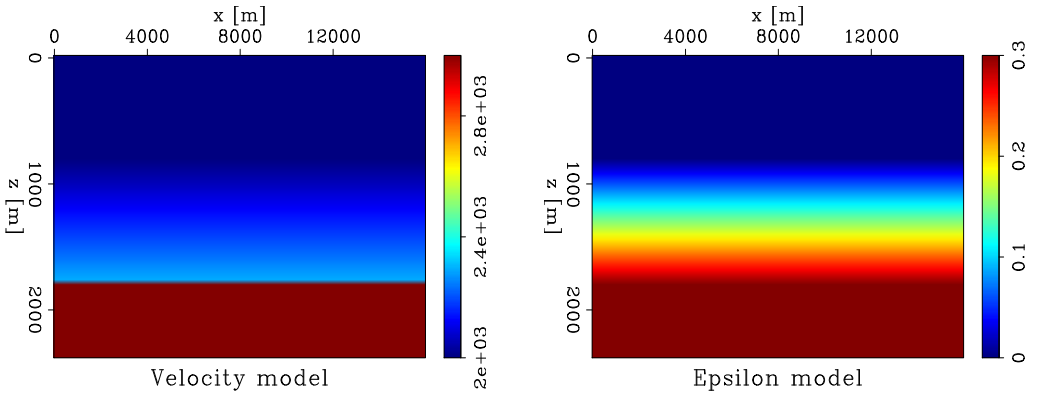
NUMERICAL TESTS

To validate my assumptions, I performed two tests. I computed the images I_{v_p} and I_ϵ using the correct v_p and δ models but varying the ϵ model. With the computed images, I evaluated the associated DSO and PSM objective functions values obtained with the different epsilon models.

Test 1

The velocity and epsilon models for the first test are shown in Figure 2. The first layer is isotropic with a vertical velocity of 2,000 meters per second (m/s). The second layer is characterized by a gradient (increasing with depth) in both the velocity and epsilon model starting at a depth of 800 meters (m). The initial values for the gradients in velocity and epsilon are, respectively, 2,000 m/s and 0. The final values (at $z = 1,800$ m) are 2,400 m/s for the velocity gradient and 0.3 for the parameter epsilon. A sharp contrast is inserted in the velocity model at a depth of 1,800 m while the epsilon model is constant below that depth. Using these models, I generated a synthetic dataset with a finite difference

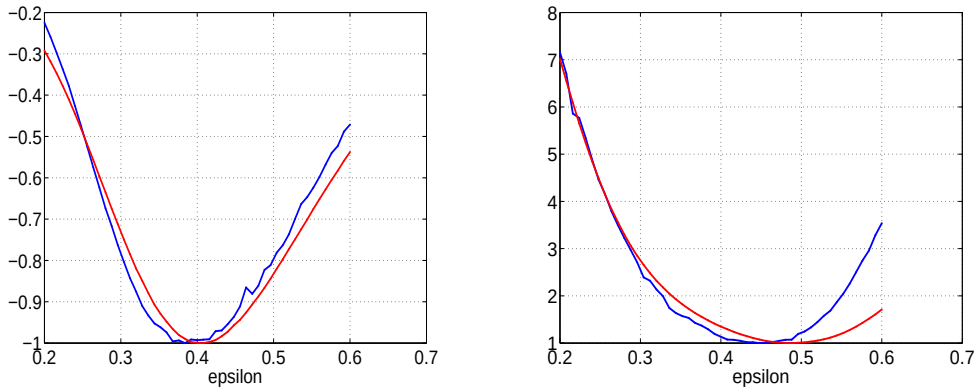
code based on the two-way wave-equation presented in the first section. The acquisition geometry is constituted of 61 sources that go from $x = 2,000$ m to $x = 14,000$ m at depth zero and receivers everywhere on the same surface as the sources. Receiver spacing is 20 m.



(a) (b)

Figure 2: Test 1. Vertical velocity (a) and epsilon (b) models used for the generation of the synthetic dataset. The initial values for the gradients in velocity and epsilon are, respectively, $2,000$ m/s and 0. The final values (at $z = 1,800$ m) are $2,400$ m/s for the velocity gradient and 0.3 for the parameter epsilon. [**ER**] carlo1/. test1-velocity,test1-epsilon

I generated different epsilon models by rescaling the gradient in the second layer: The initial value is always 0, while the final one (at $z = 1,800$ m) goes from 0.2 to 0.6. I used the different epsilon models to compute I_{v_p} and I_ϵ.



(a) (b)

Figure 3: Test 1. PSM (a) and DSO (b) curves computed using I_{v_p} (red) and I_ϵ (blue) with different epsilon models. The DSO curve associated to I_ϵ seems to indicate a greater sensitivity to the variations in the epsilon model. [**NR**] carlo1/. test1-psm1eps,test1-dso1eps

The red curves in the graphs of Figure 3 show the values of the PSM (a) and DSO (b) objective function with respect to the epsilon model computed with I_{v_p}. The blue curves are the ones computed using I_ϵ. The curves obtained using the two different images are almost equivalent for the case of the PSM objective function. For the case of the DSO, though, the one associated to I_ϵ seems to indicate a greater sensitivity to the variations in the epsilon

model. Indeed, especially around the minimum location, the DSO curve associated with I_ϵ is steeper than the one associated with I_{v_p}, meaning that the same variation in the ϵ model produces a bigger change in I_ϵ rather than in I_{v_p}.

Test 2

The velocity and epsilon models for the second test are shown in Figure 4. The first layer is characterized by a constant vertical velocity of 2,000 m/s. The epsilon model is composed of a constant background ($\epsilon = 0$) with a Gaussian anomaly centered at $z = 1,200$ m and $x = 8,000$ m. The anomaly has a maximum value of 0.3. A sharp contrast that simulates a tilted reflector is inserted in both the velocity and epsilon models. The dip of the reflector is 5 degrees. These models were used to generate a synthetic dataset with the same acquisition geometry of Test 1.

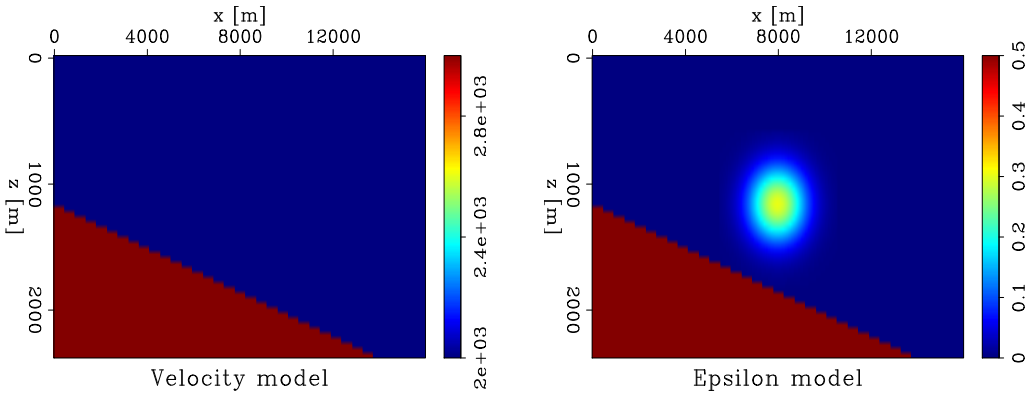
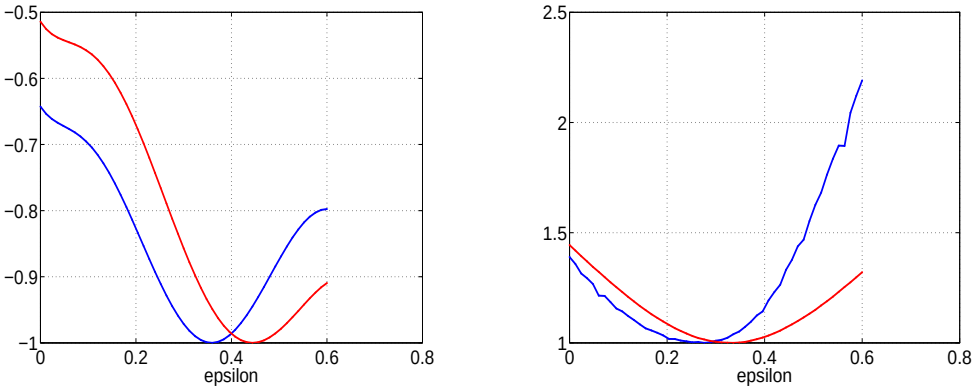


Figure 4: Test 2. Vertical velocity (a) and epsilon (b) models used for the generation of the synthetic dataset. [**ER**] carlo1/. test2-velocity,test2-epsilon

I generated the different epsilon models by rescaling the epsilon anomaly in the first layer. The shape of the anomaly is always the same, but the maximum value (at $x = 8,000$ m, $z = 1,200$ m) varies between 0 and 0.7. As for Test 1, I used the different epsilon models to compute I_{v_p} and I_ϵ.

The red curves in the graphs of Figure 5 show the values of the PSM (a) and DSO (b) objective function with respect to the epsilon model computed with I_{v_p}. The blue curves are the ones relative to I_ϵ. The curves computed with the two different images (I_{v_p} and I_ϵ) show more significant differences with respect to those of Test 1, probably because the fact that the dipping reflector enhance the contribution of the events associated with large reflection angles. These events are indeed the ones that carries most of the information about the parameter ϵ. Also in this case, the variations in the epsilon model seem to affect I_ϵ more than I_{v_p}.

The minimum/maximum of the curves in both tests do not coincide with the model used to generate the data (the minimum should be at $\epsilon = 0.3$). I attribute this to finite-difference acquisition artifacts. Moreover, the location of the minimum is different for the curves



(a) (b)

Figure 5: Test 2. PSM (a) and DSO (b) curves computed using I_{v_p} (red) and I_ϵ (blue) with different epsilon models. Notice that the curves do not have their minimum at the same locations. The DSO curve associated to I_ϵ seems to indicate a greater sensitivity to the variations in the epsilon model. [**NR**] carlo1/. test2-psm1eps,test2-dso1eps

computed using I_{v_p} and I_ϵ and seems to be closer to the correct position when using the latter. More studies are needed to correctly analyze the problem.

CONCLUSIONS

In this paper, I presented a workflow for WEMVA derived using the VTI approximation. I used a pseudo-acoustic two-way wave-equation as the modeling engine. I also presented some preliminary results on the use of the generalized images as input for WEMVA. Although the preliminary tests seem to suggest that the proposed procedure could improve the accuracy and convergence rate of the velocity analysis, it is still unclear if the use of generalized images as input for WEMVA produces significant benefits with respect to the conventional approach.

FUTURE WORK

Additional work is needed to both evaluate the proposed WEMVA workflow based on the pseudo-acoustic two-way wave-equation and to further investigate the suggested approach of using the generalized images. For the first task, a code for the complete tomographic loop is needed. For the evaluation of the new proposed approach, on the other hand, further theoretical studies are needed for both gaining a better understanding of the problem and designing additional tests.

ACKNOWLEDGMENT

I would like to thank my advisor Giuseppe Drufuca, Nicola Bienati and Biondo Biondi for giving me the opportunity to spend some time working in the SEP group. I would like to thank Xukai Shen and Elita Li for helping me and letting me use their codes. Ali Almomin

and Guillaume Barnier greatly helped me with useful ideas and suggestions. I would also thank all the people of the Stanford Exploration Project research group for their support and help.

REFERENCES

Alkhalifah, T., 1998, Acoustic approximations for processing in transversely isotropic media: Geophysics, **63**, 623–631.

Duveneck, E., P. Milcik, M. P. Bakker, and C. Perkins, 2008, Acoustic VTI wave equations and their application for anisotropic reverse-time migration: SEG Expanded Abstracts, **27**.

Fletcher, R., X. Du, and J. P. Fowler, 2009, Stabilizing acoustic reverse-time migration in TTI media: SEG Expanded Abstracts, **28**.

Grechka, V., L. Zhan, and J. W. Rector, 2004, Shear waves in acoustic anisotropic media: Geophysics, **69**, 576–582.

Li, Y. and B. Biondi, 2011, Migration velocity analysis for anisotropic models: SEG Expanded Abstracts.

Li, Y., P. Shen, and C. Perkins, 2012, Vti migration velocity analysis using RTM: SEP-Report, **147**, 57–72.

Plessix, E. R., 2006, A review of the adjoint-state method for computing the gradient of a funtional with geophysical applications: Geophys. J. Int., **167**, 495–503.

Sava, P. and S. Fomel, 2006, Generalized imaging conditions for wave equation migration: CWP report, 524.

Shen, P., 2004, Wave equation migration velocity analysis by differential semblance optimization: PhD thesis, Rice University.

Symes, W. W. and M. Kern, 1994, Inversion of reflection seismograms by differential semblance analysis: Algorithm structure and synthetic examples: Geophysics, **42**, 565–614.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

Thomsen, A. L., 1986, Weak elastic anisotropy: Geophysics, **51**, 1954–1966.

Toldi, L. J., 1989, Velocity analysis without picking: Geophysics, **54**, 191–199.

Weibull, W. W. and B. Arntsen, 2014, Anisotropic migration velocity analysis using reverse-time migration: Geophysics, **79**.

Zhang, Y. and H. Zhang, 2009, A stable TTI reverse time migration and its implementation: SEG Expanded Abstracts, **28**, 3–11.

Multi-model full-waveform inversion

Musa Maharramov and Biondo Biondi

ABSTRACT

We propose a multi-model formulation of full-waveform inversion that is similar to image decomposition into a "cartoon" and "texture" used in image processing. Inversion problem is formulated as unconstrained multi-norm optimization that can be solved using conventional iterative solvers. We demonstrate the proposed model decomposition approach by recovering a blocky subsurface seismic model from noisy data in time-lapse and single-model full-waveform inversion problems.

INTRODUCTION

Maharramov and Biondi (2014b) proposed a total-variation (TV) regularization technique for robust recovery of production-induced changes in the subsurface velocity model. Maharramov and Biondi (2014a) proposed a TV-regularized constrained full-waveform inversion (CFWI) technique that achieves better constraining of the subsurface model in zones of poor illumination. In both cases the recovered model was assumed to be "blocky"—i.e., contain areas of small (and predominantly monotone) velocity variation, as well as a few sharp contrasts along geologic interfaces. While sensible in many applications, "blocky" models represent an oversimplification of true physical properties. Even when the assumption of blocky behavior is justified, fitting noisy data may still result in an oscillatory model.

In this work we investigate splitting of a subsurface model into the sum of a "blocky" and "wiggly" components, and formulate a full-waveform inversion problem for recovering both components. Conceptually, our approach is similar to image decomposition into a "cartoon" and "texture" (Meyer, 2001), where "texture" is defined as a highly oscillatory pattern. We describe an implementation of multi-model full-waveform inversion of time-lapse and single-acquisition datasets, and demonstrate the method on two synthetic examples.

METHOD

A multi-model full-waveform inversion can be posed as an unconstrained regularized multiple-norm optimization problem:

$$\begin{aligned} \min_{\mathbf{m},\mathbf{m}_b} \; & \|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2 \; + \; \alpha \||\nabla \mathbf{m}_b|\|_1 + \beta \|\Delta \mathbf{m}_w\|_2^2, \\ & \mathbf{m} \; = \; \mathbf{m}_b + \mathbf{m}_w, \end{aligned} \tag{1}$$

where $\mathbf{m}_{b,w}$ are two *uncorrelated* "blocky" and "wiggly" components of the model $\mathbf{m}$, with $\|\mathbf{m}_2\|_2 \ll \|\mathbf{m}_1\|_2$. The ℓ_1 norm of the gradient, or the *total variation* seminorm (Triebel, 2006), favors sharp contrasts over oscillations in $\mathbf{m}_b$, while the ℓ_2 norm favors small oscillations over large contrasts. Solving (1) produces a model split into two components, with

one component exhibiting mostly blocky, and the other—oscillatory behavior. However, clean separation cannot be realistically achieved. Even in the simplest case of image TV denoising with the trivial modeling operator $\mathbf{F}(\mathbf{m}) \equiv \mathbf{m}$, the two recovered components remain correlated (Meyer, 2001; Osher et al., 2005).

A problem analogous to (1) can be formulated for time-lapse FWI,

$$\begin{aligned} \min_{\mathbf{m}_1,\mathbf{m}_2,\mathbf{m}_b} \quad & \|\mathbf{F}(\mathbf{m}_1) - \mathbf{d}_1\|_2^2 \; + \|\mathbf{F}(\mathbf{m}_2) - \mathbf{d}_2\|_2^2 \; + \\ & \alpha \| |\nabla \mathbf{m}_b| \|_1 + \beta \|\Delta \mathbf{m}_w\|_2^2, \\ & \mathbf{m}_2 - \mathbf{m}_1 \;=\; \mathbf{m}_b + \mathbf{m}_w, \end{aligned} \tag{2}$$

where $\mathbf{m}_{1,2}$ and $\mathbf{d}_{1,2}$ are the baseline and monitor subsurface models and recorded data.

We solve problems (1) and (2) by applying the nonlinear conjugate gradients algorithm (Nocedal and Wright, 2006) after smoothing the TV term,

$$\| |\nabla \mathbf{m}| \|_1 \;\approx\; \| \sqrt{|\nabla_{\mathbf{x}} m|^2 + \epsilon} \|_1, \tag{3}$$

where $\epsilon \approx 10^{-5}$ is chosen as a threshold for realistic values of the slowness. Regularization parameters α and β are chosen as follows. The value of α is chosen as in the standard TV-regularization problem with $\mathbf{m} = \mathbf{m}_b$ ($\mathbf{m}_2 - \mathbf{m}_1 = \mathbf{m}_b$ for time-lapse FWI) used in (Maharramov and Biondi, 2014b). Then β is chosen sufficiently large to make oscillatory component close to zero, and gradually reduced until the models become correlated.

Alternative constrained optimization problems can be formulated instead of (1) and (2), and solved using the approach described by Maharramov and Biondi (2014a). Alternatively, *split-Bregman* of Goldstein and Osher (2009) can be applied directly to (1) and (2).

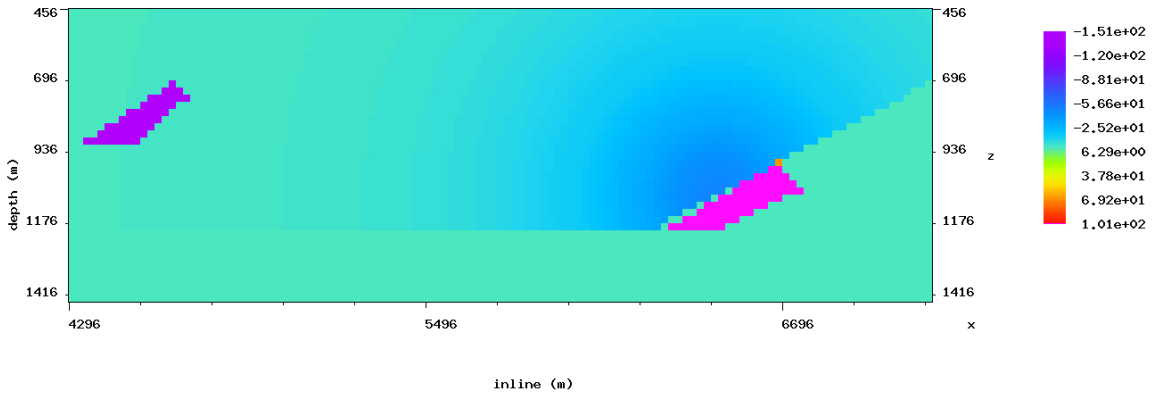


Figure 1: True model difference showing two blocky anomalies (-150 m/s and 100 m/s) and a smooth velocity change peaking at -50 m/s in the overburden above the right anomaly. [**CR**] musa3/. truediff

NUMERICAL EXAMPLES

We demonstrate multi-model inversion on the 7dB SNR Marmousi synthetic that we used in (Maharramov and Biondi, 2014b,a). See these papers for the details of numerical imple-

mentation.

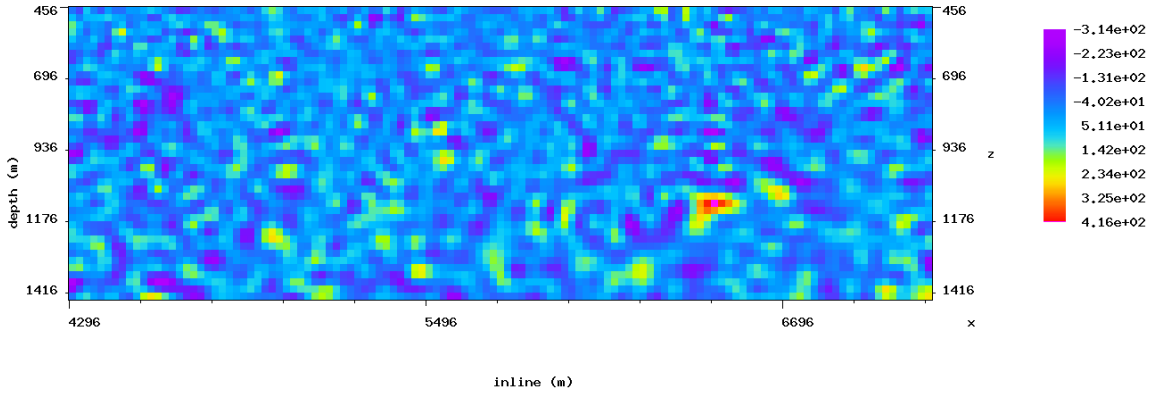


Figure 2: Model difference reconstructed using parallel difference algorithm (Asnaashari et al., 2012) for the 7dB SNR synthetic. Both amplitudes and locations of the anomalies are poorly resolved. [**CR**] musa3/. pardiff

For the time lapse example, a true model was generated consisting of "blocky" anomalies of -150 m/s and 100 m/s and a smooth velocity variation peaking at -50 m/s above the right anomaly (see Fig 4).

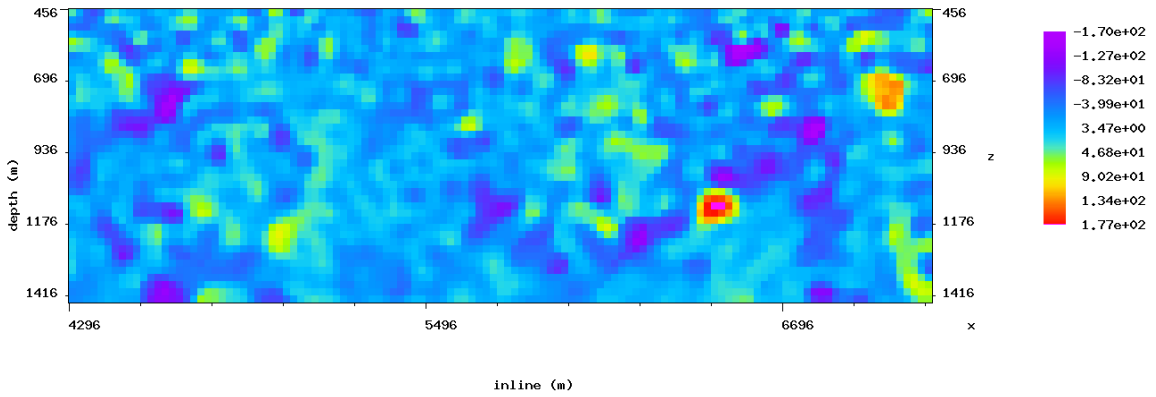


Figure 3: Blocky component of the model difference recovered by solving (2). The anomalies are resolved better, with amplitudes close to true values. The decomposition reveals partial recovery of the negative smooth velocity change over the right anomaly, however, most of the smooth velocity change ended up in the oscillatory component in Figure 4. [**CR**] musa3/. blocky

Note that the position of the smooth gradient and the small magnitude of velocity perturbations ($< 5\%$ of baseline), in combination with noisy data, make resolution of the model difference very challenging in this case.

This is confirmed by the result of parallel difference algorithm (Asnaashari et al., 2012) in Figure 2. The anomalies are hard to identify, and their amplitudes are overestimated.

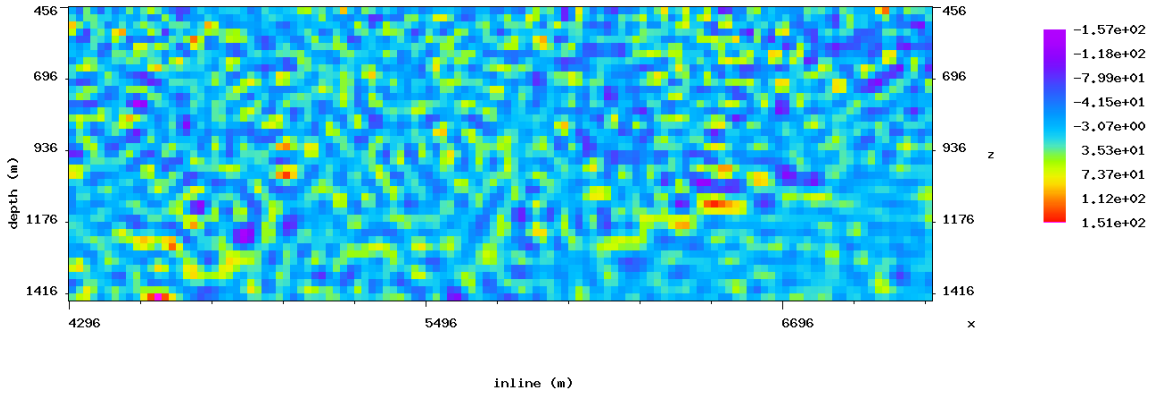


Figure 4: Oscillatory component of the model difference recovered by solving (2). Note that the two components appear to be mostly uncorrelated. [**CR**] musa3/. wiggly

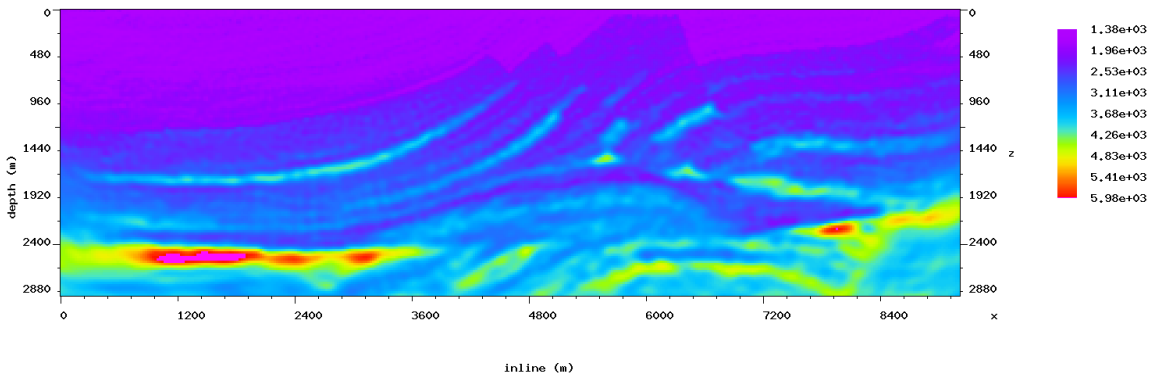


Figure 5: Blocky component of the baseline model recovered by solving (2) for the 7dB SNR synthetic. [**CR**] musa3/. n4blocky

The blocky component in Figure 3 was recovered by solving (2) with $\alpha = 10^{-6}$ and $\beta = 10^{-4}$. It appears to be a better approximation of the true model difference, both qualitatively and quantitatively. Note the partial recovery of the smooth velocity change in the overburden.

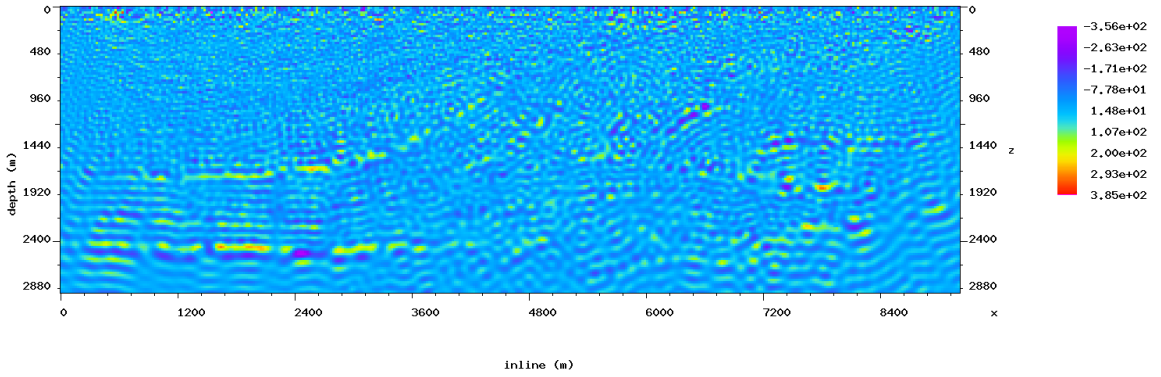


Figure 6: Oscillatory component of the baseline model recovered by solving (2) for the 7dB SNR synthetic. The model is largely uncorrelated with the blocky model of Figure 3, however, some of the sharp contrasts lave leaked into the wiggly model. [**CR**] musa3/. n4wiggly

The oscillatory component shown in Figure 4 is weakly correlated with the blocky model, and mostly represents noise removed from the model difference by he multiscale inversion (2). Note, however, that the oscillatory component may contain some of the smooth overburden velocity change—see the negative velocity zone above the location of right anomaly in Figure 4. A hierarchical multiscale decomposition approach similar to (Tadmor et al., 2004) may be applied to further decompose the model into blocky, *smoothly varying* and oscillatory components.

Figures 5 and 6 show the result of solving (1) with the 7dB SNR synthetic. The two components demonstrate the typical pattern with multiscale multinorm decompositions: while the two blocky and oscillatory models are largely uncorrelated, some of the sharp contrasts have leaked into the wiggly model in Figure 6.

CONCLUSIONS

Multi-model regularization of full-waveform inversion can be used for automated multiscale model decomposition. This can be exploited for isolating the effects of different physical processes acting on different scales, or separating useful information from the effects of fitting noisy data. The multi-model FWI can be implemented using the existing nonlinear unconstrained iterative solver frameworks with modest computational overhead, however, comparison with alternative methods (Goldstein and Osher, 2009; Osher et al., 2005; Cai et al., 2010; Boyd et al., 2011) is necessary. Hierarchical model decompositions (Tadmor et al., 2004) and application to field data will be the subject of future work.

ACKNOWLEDGMENTS

The authors would like to thank Jon Claerbout and Stewart Levin for a number of useful discussions, and the Stanford Center for Computational Earth and Environmental Sciences for providing computing resources.

REFERENCES

Asnaashari, A., R. Brossier, S. Garambois, F. Audebert, P. Thore, and J. Virieux, 2012, Time-lapse imaging using regularized FWI: A robustness study: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi:10.1190/segam2012-0699.1, 1–5.

Boyd, S., N. Parikh, E. Chu, B. Peleato, and J. Eckstein, 2011, Distributed optimization and statistical learning via the alternating direction method of multipliers: Foundations and Trends in Machine Learning, **3**, 1–122.

Cai, J., S. Osher, and Z. Shen, 2010, Split Bregman methods and frame based image restoration: Multiscale Modeling & Simulation, **8**, 337–369.

Goldstein, T. and S. Osher, 2009, The Split Bregman method for l1-regularized problems: SIAM Journal on Imaging Sciences, **2**, 323–343.

Maharramov, M. and B. Biondi, 2014a, Improved depth imaging by constrained full-waveform inversion: SEP Report, **155**, 19–24.

———, 2014b, Robust joint full-waveform inversion of time-lapse seismic datasets with total-variation regularization: SEP Report, **155**, 199–208.

Meyer, Y., 2001, Oscillating patterns in image processing and nonlinear evolution equations: AMS.

Nocedal, J. and S. J. Wright, 2006, Numerical optimization: Springer.

Osher, S., M. Burger, D. Goldfarb, J. Xu, and W. Yin, 2005, An iterative regularization method for total variation-based image restoration: Multiscale Modeling & Simulation, **4**, 460–489.

Tadmor, E., S. Nezzar, and L. Vese, 2004, A multiscale image representation using hierarchical (bv, l 2) decompositions: Multiscale Modeling & Simulation, **2**, 554–579.

Triebel, H., 2006, Theory of function spaces: Birkh auser.

Reformulating TFWI

Ali Almomin

ABSTRACT

Tomographic full waveform inversion (TFWI) provides a robust but expensive method to invert the seismic data. Scale separation of the model greatly reduces the cost but adds complexity to theory and the implementation of the inversion. In this paper, I provide two approaches that reduce the complexity of TFWI. First, I rederive TFWI with one model only in an abstract formulation that is applicable to any form of wave-equation. Then, I provide a new approximation to the inversion that can potentially provide more accurate results.

INTRODUCTION

Previously, we reduced the cost of TFWI by separating the extended model into two components: a non-extended smooth background and an extended rough perturbation (Biondi and Almomin, 2012; Almomin and Biondi, 2013). This might have caused some confusion on the resulting relationship and balance between these two parameters and their relationship to the original model. Furthermore, the interpretation of these two parameters limited the way we could separate them and increased the difficulty of moving to different wave-equations, such as the elastic. To overcome these limitations, I first derive the "original" TFWI in an abstract way that makes it applicable to different forms of the wave-equation. Then, I derive TFWI using the two-parameter approach. Next, I rederive TFWI while keeping one abstract model. Finally, I provide an alternative method to derive TFWI that is potentially more accurate than the previous TFWI and closer to the original TFWI.

THEORY

To solve any nonlinear inversion problem in a gradient-based method, we only need to evaluate two quantities: the objective function and the gradient. Evaluating the objective function requires the forward modeling operator whereas evaluating the gradient requires the linearized adjoint operator, i.e., the adjoint of the linearized forward operator. The linearized forward operator can be computed simply by taking the derivative of the forward modeling operator with respect to the model space. This linearized operator relates a (preferably small) perturbation of the model space to a perturbation in the data space. In the wave-equation, the effect of linearized operator is an explicit scattering of the waves by the model perturbations or an imaging these perturbations by the adjoint operator. I will refer to this type of scattering as linearization scattering. For the purposes of this report, I will ignore the regularization term of TFWI and only focus on the data fitting term. Moreover, a tilde ($\sim$) above a parameter or operator indicates that it is extended.

TFWI

First, we look at TFWI without any approximations or shortcuts, which I refer to as the original TFWI. The original TFWI objective function is:

$$J_{\text{TFWI}}(\widetilde{\mathbf{m}}) = \frac{1}{2}\|\widetilde{\mathbf{f}}(\widetilde{\mathbf{m}}) - \mathbf{d}_{\text{obs}}\|_2^2 = \frac{1}{2}\|\mathbf{r}(\widetilde{\mathbf{m}})\|_2^2, \tag{1}$$

where $\widetilde{\mathbf{m}}$ is the extended model, $\widetilde{\mathbf{f}}$ is the extended forward modeling operator, $\mathbf{d}_{\text{obs}}$ is the observed surface data, and $\mathbf{r}$ is the residual. In the extended modeling operator, all parts of the extended model interact with the wavefield in every propagation step, which is the main source of cost increase in computation. The gradient of the objective function is:

$$\frac{\partial J_{\text{TFWI}}(\widetilde{\mathbf{m}})}{\partial \widetilde{\mathbf{m}}} = \left(\frac{\partial \mathbf{r}(\widetilde{\mathbf{m}})}{\partial \widetilde{\mathbf{m}}}\right)^* \mathbf{r}(\widetilde{\mathbf{m}}). \tag{2}$$

The derivative of the residual is computed as:

$$\frac{\partial \mathbf{r}(\widetilde{\mathbf{m}})}{\partial \widetilde{\mathbf{m}}} = \frac{\partial \widetilde{\mathbf{f}}(\widetilde{\mathbf{m}})}{\partial \widetilde{\mathbf{m}}} = \widetilde{\mathbf{L}}(\widetilde{\mathbf{m}}), \tag{3}$$

where $\widetilde{\mathbf{L}}$ is the extended linearized modeling operator. I can now compute the gradient as follows:

$$\mathbf{g}(\widetilde{\mathbf{m}}) = \widetilde{\mathbf{L}}^*(\widetilde{\mathbf{m}})\mathbf{r}(\widetilde{\mathbf{m}}). \tag{4}$$

Although the previous equations are simple and direct, they can be very expensive and difficult to compute, depending on the extension axes of choice. Therefore, we need to find a cheaper approach to approximating these computations.

Two-parameter approximation

The previously published approximation to TFWI utilizes the two-parameter approximation that separates the extended model into two components:

$$\widetilde{\mathbf{m}} \approx \mathbf{b} + \widetilde{\mathbf{p}}, \tag{5}$$

where $\mathbf{b}$ is the non-extended smooth background component and $\widetilde{\mathbf{p}}$ an extended rough perturbation component. The underlying assumption is that the extended portion will not have a smooth component. It is also easy to understand conceptually: we are separating the parts of the model that affect the transmission (or forward scattering) from the parts that affect the reflection (or back scattering). Now, I can approximate the modeling operator using Taylor's expansion as follows:

$$\widetilde{\mathbf{f}}(\widetilde{\mathbf{m}}) \approx \widetilde{\mathbf{f}}(\mathbf{b} + \widetilde{\mathbf{p}}) \approx \mathbf{f}(\mathbf{b}) + \widetilde{\mathbf{L}}(\mathbf{b})\widetilde{\mathbf{p}}. \tag{6}$$

This approximate modeling operator simply adds the data resulting from propagating the wavefield with the smooth component of the model (the $\mathbf{f}(\mathbf{b})$ term) to the data resulting from scattering the wavefield with the rough component of the model (the $\widetilde{\mathbf{L}}(\mathbf{b})\widetilde{\mathbf{p}}$ term). It is important to notice that this scattering is a result of the approximation we used by model separation, which is similar to but independent from the linearization scattering

due to gradient calculation. I will refer to this second type of scattering as approximation scattering. Both the linearization scattering and approximation scattering use Taylor's expansion, which might make things a bit confusing. Moreover, the scattering with the rough component happens only once, so multiples cannot be modeled with this operator unless the rough component was added to the smooth background. Using the approximate operator, the residuals can be calculated as:

$$\mathbf{r}(\mathbf{b},\widetilde{\mathbf{p}}) = \mathbf{f}(\mathbf{b}) + \widetilde{\mathbf{L}}(\mathbf{b})\widetilde{\mathbf{p}} - \mathbf{d}_{\text{obs}}. \tag{7}$$

Next, I calculate the residual derivative with respect to the perturbation as:

$$\frac{\partial \mathbf{r}(\mathbf{b},\widetilde{\mathbf{p}})}{\partial \widetilde{\mathbf{p}}} = \widetilde{\mathbf{L}}(\mathbf{b}), \tag{8}$$

and the residual derivative derivative with respect to the background as:

$$\frac{\partial \mathbf{r}(\mathbf{b},\widetilde{\mathbf{p}})}{\partial \mathbf{b}} = \frac{\partial \mathbf{f}(\mathbf{b})}{\partial \mathbf{b}} + \frac{\partial \widetilde{\mathbf{L}}(\mathbf{b})}{\partial \mathbf{b}}\widetilde{\mathbf{p}} = \mathbf{L}(\mathbf{b}) + \mathbf{T}(\mathbf{b},\widetilde{\mathbf{p}}), \tag{9}$$

where $\mathbf{T}$ is the tomographic operator that is a function of two variables: a propagation variable and a scattering variable. This tomographic operator has two scattering terms because we first applied the approximation linearization, and then perturbed that model using the linearization scattering, which is effectively a scattering of a scatterer. In other words, it is the second derivative of the modeling operator with respect to the model. Finally, I can calculate the gradient of $\widetilde{\mathbf{p}}$ as:

$$\mathbf{g}_{\widetilde{\mathbf{p}}}(\mathbf{b},\widetilde{\mathbf{p}}) = \widetilde{\mathbf{L}}^*(\mathbf{b})\mathbf{r}(\mathbf{b},\widetilde{\mathbf{p}}), \tag{10}$$

and the gradient of $\mathbf{b}$ as:

$$\mathbf{g}_{\mathbf{b}}(\mathbf{b},\widetilde{\mathbf{p}}) = \mathbf{L}^*(\mathbf{b})\mathbf{r}(\mathbf{b},\widetilde{\mathbf{p}}) + \mathbf{T}^*(\mathbf{b},\widetilde{\mathbf{p}})\mathbf{r}(\mathbf{b},\widetilde{\mathbf{p}}). \tag{11}$$

One issue with the gradient of $\mathbf{b}$ is the term $\mathbf{L}^*(\mathbf{b})\mathbf{r}(\mathbf{b},\widetilde{\mathbf{p}})$ which can potentially cycle-skip since the operator is not extended. Another issue is that multiples will not be modeled (neither by the non-extended part nor by the extended part) unless we feed the extended perturbation back into the non-extended background. To circumvent these problems, Almomin and Biondi (2013) proposed the nested inversion scheme where the objective function was separated into two components and the observed data was redefined in order to avoid this term in the gradient of $\mathbf{b}$. This nested scheme requires a balance between how many iterations we perform in each inner loop as well as a good balance on how to mix and separate the smooth and rough components. Having these components in different dimensions only adds complexity to the understanding of the problem.

One-parameter approximation

The previous approximation is based on the fact that the extended part of the model only contains rough components. Furthermore, it requires a nested inversion scheme with several steps that ensure different scales of the model are simultaneously inverting. Also, implementing the previous scheme for other wave-equations, such as the elastic wave-equation, might not be trivial in terms of which variable to extend and how to separate and mix

them. Therefore, I rederive the previous approximation in an equivalent but generalized way using one parameter and an abstract separation operator as follows:

$$\widetilde{\mathbf{m}} = \mathbf{C}\widetilde{\mathbf{m}} + [\mathbf{I} - \mathbf{C}]\widetilde{\mathbf{m}}, \tag{12}$$

where $\mathbf{C}$ is the separation operator. For the operator approximation to be accurate, I need the following inequality to hold:

$$\|\mathbf{C}\widetilde{\mathbf{m}}\| \gg \|[\mathbf{I} - \mathbf{C}]\widetilde{\mathbf{m}}\|. \tag{13}$$

Now, I can approximate the modeling operator using Taylor's expansion as follows

$$\widetilde{\mathbf{f}}(\widetilde{\mathbf{m}}) = \widetilde{\mathbf{f}}(\mathbf{C}\widetilde{\mathbf{m}} + [\mathbf{I} - \mathbf{C}]\widetilde{\mathbf{m}}) \approx \widetilde{\mathbf{f}}(\mathbf{C}\widetilde{\mathbf{m}}) + \widetilde{\mathbf{L}}(\mathbf{C}\widetilde{\mathbf{m}})[\mathbf{I} - \mathbf{C}]\widetilde{\mathbf{m}}. \tag{14}$$

For the previous equation to be efficient, the operator $\mathbf{C}$ needs to restrict the extended part of the model. From the two requirements of the operator $\mathbf{C}$, we can see that separating the model into smooth and rough components is not the only possibility. In fact, it is more optimal to separate it into a non-extended part and an exteded part, both of which can contain rough or smooth components. This latter separation will make the approximation scattering even smaller compared to the propagation term. The residual can now be written as:

$$\mathbf{r}(\widetilde{\mathbf{m}}) = \widetilde{\mathbf{f}}(\mathbf{C}\widetilde{\mathbf{m}}) + \widetilde{\mathbf{L}}(\mathbf{C}\widetilde{\mathbf{m}})[\mathbf{I} - \mathbf{C}]\widetilde{\mathbf{m}} - \mathbf{d}_{\text{obs}}. \tag{15}$$

Next, I calculate the residual derivative as follows:

$$\begin{aligned}\frac{\partial \mathbf{r}(\widetilde{\mathbf{m}})}{\partial \widetilde{\mathbf{m}}} &= \frac{\partial \widetilde{\mathbf{f}}(\mathbf{C}\widetilde{\mathbf{m}})}{\partial \widetilde{\mathbf{m}}} + \frac{\partial \widetilde{\mathbf{L}}(\mathbf{C}\widetilde{\mathbf{m}})}{\partial \widetilde{\mathbf{m}}}[\mathbf{I} - \mathbf{C}]\widetilde{\mathbf{m}} + \widetilde{\mathbf{L}}(\mathbf{C}\widetilde{\mathbf{m}})[\mathbf{I} - \mathbf{C}] \\ &= \widetilde{\mathbf{L}}(\mathbf{C}\widetilde{\mathbf{m}})\mathbf{C} + \widetilde{\mathbf{T}}(\mathbf{C}\widetilde{\mathbf{m}}, [\mathbf{I} - \mathbf{C}]\widetilde{\mathbf{m}})\mathbf{C} + \widetilde{\mathbf{L}}(\mathbf{C}\widetilde{\mathbf{m}})[\mathbf{I} - \mathbf{C}].\end{aligned} \tag{16}$$

Since I used one model parameter only, the terms can be cancelled resulting in the following residual derivative:

$$\frac{\partial \mathbf{r}(\widetilde{\mathbf{m}})}{\partial \widetilde{\mathbf{m}}} = \widetilde{\mathbf{L}}(\mathbf{C}\widetilde{\mathbf{m}}) + \widetilde{\mathbf{T}}(\mathbf{C}\widetilde{\mathbf{m}}, [\mathbf{I} - \mathbf{C}]\widetilde{\mathbf{m}})\mathbf{C}. \tag{17}$$

Finally, I can calculate the gradient as follows:

$$\widetilde{\mathbf{g}}(\widetilde{\mathbf{m}}) = \widetilde{\mathbf{L}}^*(\mathbf{C}\widetilde{\mathbf{m}})\mathbf{r}(\widetilde{\mathbf{m}}) + \mathbf{C}^*\widetilde{\mathbf{T}}^*(\mathbf{C}\widetilde{\mathbf{m}}, [\mathbf{I} - \mathbf{C}]\widetilde{\mathbf{m}})\mathbf{r}(\widetilde{\mathbf{m}}). \tag{18}$$

Alternative approximation

Now I present an alternative formulation to approximate the expensive TFWI. Instead of first approximating the modeling operator and then taking the derivative of the residual, I first take the "expensive" derivative followed approximating the calculation. This way, I can focus on approximating the propagation operator only since it is the most expensive step, instead of approximating the modeling operator. To do so, I need to rewrite our modeling operator and gradient as functions of the propagation operator. The modeling operator can be written as:

$$\widetilde{\mathbf{f}}(\widetilde{\mathbf{m}}) = \mathbf{K}^*\widetilde{\mathbf{P}}(\widetilde{\mathbf{m}})\mathbf{K}\mathbf{s}, \tag{19}$$

where $\mathbf{K}$ is a spatial padding operator, $\widetilde{\mathbf{P}}$ is the extended propagation operator, and $\mathbf{s}$ is the source. Next, I write the linearized modeling operator as:

$$\widetilde{\mathbf{L}}(\widetilde{\mathbf{m}}) = \mathbf{K}^*\widetilde{\mathbf{P}}(\widetilde{\mathbf{m}})\mathbf{D}\widetilde{\mathbf{U}}_\mathbf{s}(\widetilde{\mathbf{m}}), \tag{20}$$

where $\mathbf{D}$ is a second time derivative, and $\widetilde{\mathbf{U}}_\mathbf{s}$ is the extended scattering operator that convolve the input with a source wavefield to calculate a scattered wavefield. The source wavefield that is used in the scattering operator is calculated as:

$$\mathbf{u}_\mathbf{s}(\widetilde{\mathbf{m}}) = \widetilde{\mathbf{P}}(\widetilde{\mathbf{m}})\mathbf{K}\mathbf{s}. \tag{21}$$

I can write the adjoint modeling operator as follows:

$$\mathbf{L}^*(\widetilde{\mathbf{m}}) = \widetilde{\mathbf{U}}^*_\mathbf{s}(\widetilde{\mathbf{m}})\mathbf{D}^*\widetilde{\mathbf{P}}^*(\widetilde{\mathbf{m}})\mathbf{K}. \tag{22}$$

By examining the previous equations, we can see that the only expensive step is the propagation step. Therefore, I will approximate the propagation by a cheaper alternative, such as:

$$\widetilde{\mathbf{P}}(\widetilde{\mathbf{m}}) \approx \widetilde{\mathbf{P}}(\mathbf{C}\widetilde{\mathbf{m}}) + \widetilde{\mathbf{P}}(\mathbf{C}\widetilde{\mathbf{m}})\mathbf{D}\widetilde{\mathbf{S}}([\mathbf{I}-\mathbf{C}]\widetilde{\mathbf{m}})\widetilde{\mathbf{P}}(\mathbf{C}\widetilde{\mathbf{m}}), \tag{23}$$

where $\widetilde{\mathbf{S}}$ is a scattering operator. Finally, we use this approximate propagation operator in equations (19) to (22). Another possiblity is using a multi-scattering (recursive) propagator that can take more orders of scatterings to better approximate the extended propagation operator. Notice that the adjoint operator using the approximate propagation will not give an exact gradient of the objective function, it can be very close, depending on how well we approximated the propagation operator.

CONCLUSIONS

Scale separation of the model can add complexity to the theory and implementation of TFWI. I rederive TFWI with one model parameter in an abstract way that is applicable to any form of wave-equation. This new derivation reduces the complexity of TFWI and any ambiguity related to how different model components interact and update in the inversion. I also provide an alternative approximation to TFWI by re-ordering the gradient calculation and the Taylor's expansion of the propagation operator. The alternative approximation is easier to implement and can potentially provide more accurate results.

REFERENCES

Almomin, A. and B. Biondi, 2013, Tomographic full waveform inversion (TFWI) by successive linearizations and scale separations: SEP-Report, **149**, 51–58.

Biondi, B. and A. Almomin, 2012, Tomographic full waveform inversion (TFWI) by combining full waveform inversion with wave-equation migration velocity analysis: SEP-Report, **147**, 1–12.

Robust joint full-waveform inversion of time-lapse seismic datasets with total-variation regularization

Musa Maharramov and Biondo Biondi

ABSTRACT

We present a technique for reconstructing subsurface velocity model changes from time-lapse seismic data using full-waveform inversion (FWI). The technique is based on simultaneously inverting multiple survey vintages, with model difference regularization using the total variation (TV) seminorm. We compare the new TV-regularized time-lapse FWI with the L_2-regularized joint inversion proposed in our earlier work, using synthetic datasets that exhibit survey repeatability challenges. The results demonstrate clear advantages of the proposed TV-regularized joint inversion over alternatives methods for recovering production-induced model changes that are due to both fluid substitution and geomechanical effects.

INTRODUCTION

Effective reservoir monitoring depends on successful tracking of production-induced fluid movement in the reservoir and overburden, using input from seismic imaging, geomechanics, geology and reservoir simulation (Biondi et al., 1996). To achieve this, most traditional methods rely on the conversion of picked time shifts and reflectivity differences between migrated images into reflector movement and impedance changes. Though effective in practical applications, this approach requires a significant amount of expert interpretation and relies on quality control in the conversion process. Wave-equation image-difference tomography has been proposed as a more automatic alternative method to recover velocity changes (Albertin et al., 2006); it allows localized target-oriented inversion of model perturbations (Maharramov and Albertin, 2007). An alternative approach is based on using the high-resolution power of the full-waveform inversion (Sirgue et al., 2010) to reconstruct production-induced changes from wide-offset seismic acquisitions, and is the subject of this paper.

Time-lapse full-waveform inversion (Watanabe et al., 2004; Denli and Huang, 2009; Routh et al., 2012) is a promising technique for time-lapse seismic imaging where production-induced subsurface model changes are within the resolution of FWI. However, as with alternative time-lapse techniques, time-lapse FWI is sensitive to repeatability issues (Asnaashari et al., 2012). Non-repeatable acquisition geometries (e.g., slightly shifted source and receiver positions), acquisition gaps (e.g., due to new obstacles), different source signatures and measurement noise—all contribute to differences in the data from different survey vintages. Differences in the input datasets due to repeatability issues may easily mask valuable production-induced changes. However, even with noise-free synthetic data without any acquisition repeatability issues, numerical artifacts may contaminate the inverted difference of monitor and baseline when practical limitations are imposed on solver

iteration count. Maharramov and Biondi (2013) devised a time-lapse FWI that minimized model differences outside of areas affected by production by jointly inverting for multiple models, and imposing a regularization condition on the model difference. The joint inversion can be performed simultaneously for multiple model vintages or using an empirical technique of "cross-updating" (Maharramov and Biondi, 2013, 2014b). Maharramov and Biondi (2014b,a) applied these methods to noisy synthetic data and compared the results to alternative methods, demonstrating advantages of both target-oriented simultaneous inversion and cross-updating over alternative methods. Production-induced changes that are due to fluid substitution effects are spatially constrained to areas affected by fluid transport, while those due to geomechanical effects—e.g., stretching of the overburden—may result in smoothly varying velocity model differences. In either case, regularization terms that promote "sparsity" of spatial model-difference gradients may be expected to improve the recovery of valuable production effects while suppressing undesirable oscillatory artifacts. In this work we present the results of using total-variation (TV) regularization of the model difference in the simultaneous time-lapse FWI and demonstrate the significant improvement of the inversion results compared to alternative methods.

METHOD

Full-waveform inversion is defined as solving the following optimization problem (Tarantola, 1984; Virieux and Operto, 2009)

$$\|\mathbf{Mu}-\mathbf{d}\|_2 \;\rightarrow\; \min, \tag{1}$$

where $\mathbf{M},\mathbf{d}$ are the measurement operator and data, $\mathbf{u}$ is the solution of a forward-modeling problem

$$\mathbf{D}(\mathbf{m})\mathbf{u} \;=\; \phi, \tag{2}$$

where $\mathbf{D}$ is the forward-modeling operator that depends on a model vector $\mathbf{m}$ as a parameter, and ϕ is a source. The minimization problem (1) is solved with respect to either both the model $\mathbf{m}$ and source ϕ or just the model. In the frequency-domain formulation of the acoustic waveform inversion, the forward-modeling equation (2) becomes

$$-\omega^2 u - v^2(x^1,\ldots,x^n)\Delta u \;=\; \phi(\omega,x^1,\ldots,x^n), \tag{3}$$

where ω is a temporal frequency, n is the problem dimension, and v is the acoustic wave propagation velocity. Values of the slowness $s=1/v$ at all the points of the modeling domain constitute the model parameter vector $\mathbf{m}$. The direct problem (3) can be solved in the frequency domain, or in the time domain followed by a discrete Fourier transform in time (Virieux and Operto, 2009). The inverse problem (1) is typically solved using a multiscale approach, from low to high frequencies, supplying the output of each frequency inversion to the next step.

FWI applications in time-lapse problems seek to recover induced changes in the subsurface model using multiple datasets from different acquisition vintages. For two surveys sufficiently separated in time, we call such datasets (and the associated models) *baseline* and *monitor*.

Time-lapse FWI can be carried out by separately inverting the baseline and monitor models (*parallel difference*) ,or by inverting them sequentially with, e.g., the baseline sup-

plied as a starting model for the monitor inversion (*sequential difference*). Another alternative is to apply the *double-difference* method, with a baseline model inversion followed by a monitor inversion that solves the following optimization problem,

$$\| \left(\mathbf{M}_m^s \mathbf{u}_m - \mathbf{M}_b^s \mathbf{u}_b\right) - \left(\mathbf{M}_m \mathbf{d}_m - \mathbf{M}_b \mathbf{d}_b\right) \|_2 \ \rightarrow \ \min, \tag{4}$$

by changing the monitor model (Watanabe et al., 2004; Denli and Huang, 2009; Zheng et al., 2011; Asnaashari et al., 2012; Raknes et al., 2013). The subscripts in equation (4) denote the baseline and monitor surveys, $\mathbf{d}$ denotes the recorded data, and the $\mathbf{M}$'s are measurement operators that project the synthetic and field data onto a common grid. The superscript s indicates the measurement operators applied to the modeled data.

In all of these techniques, optimization is carried out with respect to one model at a time, albeit of different vintages at different stages of the inversion. In our method we invert for the baseline and monitor models *simultaneously* by solving either one of the following two optimization problems:

$$\alpha \|\mathbf{M}_b \mathbf{u}_b - \mathbf{d}_b\|_2^2 + \beta \|\mathbf{M}_m \mathbf{u}_m - \mathbf{d}_m\|_2^2 + \tag{5}$$
$$\gamma \| \left(\mathbf{M}_m^s \mathbf{u}_m - \mathbf{M}_b^s \mathbf{u}_b\right) - \left(\mathbf{M}_m \mathbf{d}_m - \mathbf{M}_b \mathbf{d}_b\right) \|_2^2 + \tag{6}$$
$$\alpha_1 \|\mathbf{W}_b \mathbf{R}_b (\mathbf{m}_b - \mathbf{m}_b^{\text{PRIOR}})\|_2^2 + \tag{7}$$
$$\beta_1 \|\mathbf{W}_m \mathbf{R}_m (\mathbf{m}_m - \mathbf{m}_m^{\text{PRIOR}})\|_2^2 + \tag{8}$$
$$\delta \|\mathbf{W}\mathbf{R} (\mathbf{m}_m - \mathbf{m}_b - \Delta\mathbf{m}^{\text{PRIOR}})\|_2^2 \ \rightarrow \ \min, \tag{9}$$

or

$$\alpha \|\mathbf{M}_b \mathbf{u}_b - \mathbf{d}_b\|_2^2 + \beta \|\mathbf{M}_m \mathbf{u}_m - \mathbf{d}_m\|_2^2 + \tag{10}$$
$$\gamma \| \left(\mathbf{M}_m^s \mathbf{u}_m - \mathbf{M}_b^s \mathbf{u}_b\right) - \left(\mathbf{M}_m \mathbf{d}_m - \mathbf{M}_b \mathbf{d}_b\right) \|_2^2 + \tag{11}$$
$$\alpha_1 \|\mathbf{W}_b \mathbf{R}_b (\mathbf{m}_b - \mathbf{m}_b^{\text{PRIOR}})\|_1 + \tag{12}$$
$$\beta_1 \|\mathbf{W}_m \mathbf{R}_m (\mathbf{m}_m - \mathbf{m}_m^{\text{PRIOR}})\|_1 + \tag{13}$$
$$\delta \|\mathbf{W}\mathbf{R} (\mathbf{m}_m - \mathbf{m}_b - \Delta\mathbf{m}^{\text{PRIOR}})\|_1 \ \rightarrow \ \min, \tag{14}$$

with respect to both the baseline and monitor models $\mathbf{m}_b$ and $\mathbf{m}_m$. Problem (5-9) describes time-lapse FWI with L_2 regularization of the individual models (7,8) and model difference (9) (Maharramov and Biondi, 2014b). In this work we study the second formulation (10-14) that involves L_1-regularization of the individual models and their difference. The terms (10) correspond to separate baseline and monitor inversions, the term (11) is the optional double difference term, the terms (12) and (13) are optional separate baseline and monitor inversion regularization terms (Aster et al., 2012), and the term (14) represents regularization of the model difference. In (12)-(14), $\mathbf{R}$ and $\mathbf{W}$ denote regularization and weighting operators respectively, with the subscript denoting the survey vintage where applicable. If $\mathbf{R}$ is the *gradient magnitude* operator

$$\mathbf{R} f(x,y,z) \ = \ \sqrt{f_x^2 + f_y^2 + f_z^2}, \tag{15}$$

then (12-14) become *total-variation (TV) seminorms* (Triebel, 2006). The latter case is of particular interest in this work as the minimization of the L_1 norm of gradient may promote "blockiness" of the model-difference, potentially reducing oscillatory artifacts (Rudin et al., 1992; Aster et al., 2012).

A joint inversion approach has been applied earlier to the linearized waveform inversion (Ayeni and Biondi, 2012). In Maharramov and Biondi (2013, 2014b,a), a simultaneous full-waveform inversion problem (5,9) was studied with a single model difference L_2 regularization term (9).

An implementation of the proposed simultaneous inversion algorithm requires solving a nonlinear optimization problem with twice the data and model dimensions of problems (1) and (4). The model difference regularization weights $\mathbf{W}$ and, optionally, the prior $\Delta\mathbf{m}^{\text{PRIOR}}$ may be obtained from prior geomechanical information. For example, a rough estimate of production-induced velocity changes can be obtained from time shifts (Hatchell and Bourne, 2005; Barkved and Kristiansen, 2005) and used to map subsurface regions of expected production-induced perturbation, and optionally provide a difference prior. However, successfully solving the L_1-regularized problem (10-11) is less sensitive to choice of the weighting operator $\mathbf{W}$. For example, we show below that the TV-regularization using (15) with $\mathbf{W} = 1$ recovers non-oscillatory components of the model difference, while the L_2 approach would result in either smoothing or uniform reduction of the model difference.

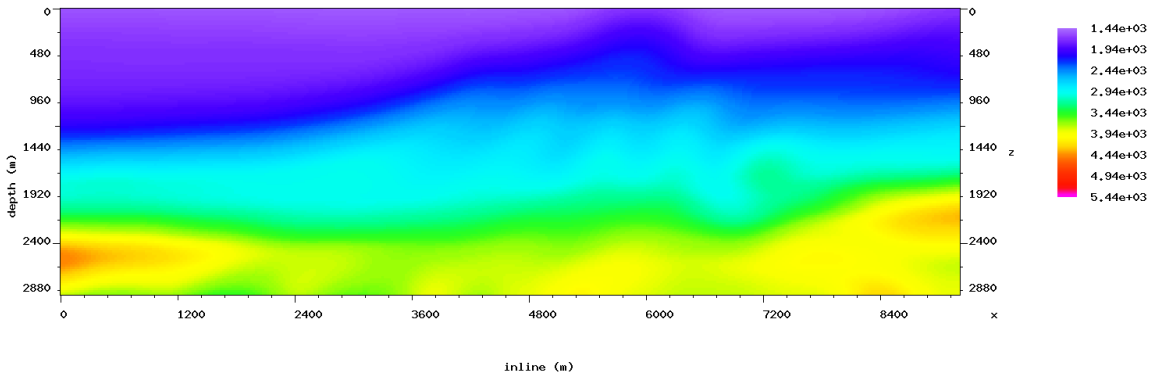


Figure 1: Starting model used in the inversion. [**CR**] musa1/. starting

In addition to the fully simultaneous inversion, Maharramov and Biondi (2013, 2014b) proposed and tested a *cross-updating* technique that offers a simple but remarkably effective approximation to minimizing the objective function (5),(9), while obviating the difference regularization and weighting operators $\mathbf{R}$ and $\mathbf{W}$ for problem (5,9). This technique consists of one standard run of the sequential difference algorithm, followed by a second run with the inverted monitor model supplied as the starting model for the second baseline inversion

$$\begin{aligned}\mathbf{m}_{\text{INIT}} \rightarrow &\text{baseline inversion} \rightarrow \text{monitor inversion} \rightarrow \\ &\text{baseline inversion} \rightarrow \text{monitor inversion},\end{aligned} \tag{16}$$

and computing the difference of the latest inverted monitor and baseline models. Process (16) can be considered as an approximation to minimizing (5) and (9) because non-repeatable footprints of both inversions are propagated to both models, canceling out in the difference. Both the simultaneous inversion and cross-updating minimize the model difference by tackling model artifacts that are in the null space of the Fréchet derivative of the forward modeling operatorss. The joint inversion minimizes the effect of such artifacts on the model difference by either minimizing the model difference term (9) in the simultaneous inversion, or by propagating these artifacts to both models in cross-updating (16).

Note that this process is not guaranteed to improve the results of the baseline and monitor model inversions but was only proposed for improving the model difference. Maharramov and Biondi (2014b,a) demonstrated a significant improvement of model difference recovery by both the L_2-regularized target-oriented simultaneous inversion and cross-updating compared to the parallel, sequential and double difference techniques. The simultaneous inversion and cross-updating yielded qualitatively similar results within the inversion target.

Here we compare joint simultaneous inversion with a TV-regularized model difference (10,14,15) to parallel difference and cross-updating.

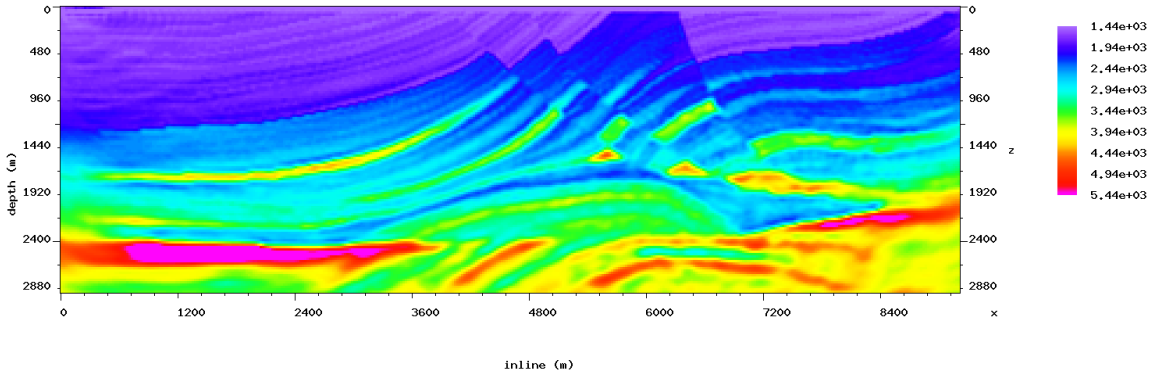


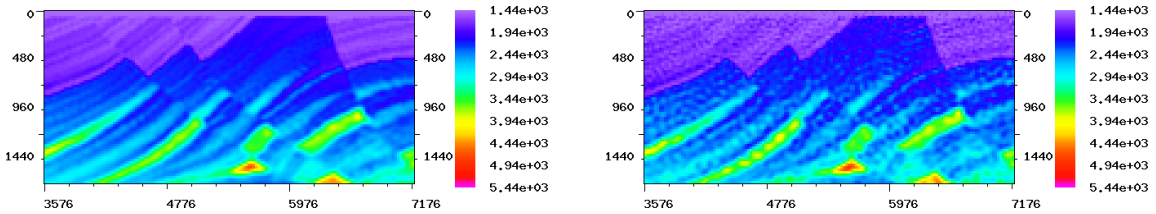
Figure 2: Baseline model inverted from noise-free synthetic data. [**CR**] musa1/. bb155Hz

NUMERICAL EXAMPLES

The Marmousi velocity model is used as a baseline, over a 384×122 grid with a 24 m grid spacing. Production-induced velocity changes are modeled as a negative −150 m/s perturbation at about 4.5 km inline 800 m depth, and a positive 200 m/s perturbation at 6.5 km inline, 1 km depth. Additionally, a smoothly varying negative velocity change, peaking at −50 m/s, was included above the positive anomaly as shown in Figure 4. The whole Marmousi model is inverted, however, only model differences for the section between the approximate inline coordinates 4 km and 6.7 km to the depth of approximately 1.4 km are shown here. The inversion is carried out in the frequency domain for 3.0, 3.6, 4.3, 5.1, 6.2, 7.5, 9.0, 10.8, 12.8, and 15.5 Hz, with the forward-modeling carried out in the time domain (Sirgue et al., 2010). Inversion frequencies are chosen based on the estimated offset to depth range of the data (Sirgue and Pratt, 2004). The baseline acquisition has 192 shots at a depth of 16 m with a 48 m spacing, and 381 receivers at a depth of 15 m with a 24 m spacing. The minimum offset is 48 m. The source function is a Ricker wavelet centered at 10.1 Hz. Absorbing boundary conditions are applied along the entire model boundary, including the surface (thus suppressing multiples). A smoothed true model (Figure 1) is used as a starting model for the initial baseline inversion (and for the initial monitor inversion in the parallel difference). The smoothing is performed using a triangular filter with a 20-sample half-window in both vertical and horizontal directions.

The result of inverting the baseline model from the clean synthetic data is shown in Figure 2. Random Gaussian noise is added to the noise-free synthetic data to produce

a noisy dataset with 7 dB signal-to-noise ratio. The noisy monitor dataset is generated for the model perturbation of Figure 4, using the same acquisition geometry and source wavelet. The results of baseline model inversion from the clean and 7 dB SNR synthetic data are shown in Figure 3(a) and Figure 3(b), respectively. Results of model difference inversion from the clean and 7 dB SNR synthetic datasets using various methods are shown in Figures 5(a),5(b),5(c) and Figures 6(a),6(b),6(c), respectively. The simultaneous inversion objective function contains only terms (10) and (14) with no difference prior, i.e., $\Delta\mathbf{m}^{\text{PRIOR}} = 0$. The model-difference regularization weights $\mathbf{W}$ in (14) are set to 1 *everywhere* in the modeling domain. The two terms in (10) are of the same magnitude and therefore α and β are set to 1. Parameter δ is set to 10^{-5} but can be varied for different acquisition source and geometry parameters. The result of the initial baseline inversion is supplied as a starting model for both $\mathbf{m}_b$ and $\mathbf{m}_m$ in the simultaneous inversion. In all the inversions, up to 10 iterations of the nonlinear conjugate gradients algorithm (Nocedal and Wright, 2006) are performed for each frequency. Neither regularization nor model priors are used in single-model inversions (i.e., in the cross-updating and parallel difference methods). Maharramov and Biondi (2014b) demonstrated significant improvement by cross-updating compared to sequential differencing, and rough qualitative equivalence of cross-updating and the L_2-regularized simultaneous inversion. Therefore, in this work we compare the TV-regularized simultaneous inversion only against parallel differencing and cross-updating.



(a) (b)

Figure 3: (a) Target area of the baseline model inverted from clean synthetic. (b) Target area of the baseline model inverted from a 7 dB SNR synthetic. In both cases the baseline model is reconstructed reasonably well, however, errors due to noise are comparable in magnitude to production-induced effects. [**CR**] musa1/. cleanbase,4base

The results of applying cross-updating to the two datasets are shown in Figures 5(b) and 6(b), respectively. The corresponding TV-regularized simultaneous inversion results are shown in Figures 5(c) and 6(c). Since problem (1) is nonlinear, supplying the result of the highest frequency inversion back to the lowest frequency and repeating the whole inversion cycle for all frequencies may result in achieving a better data fit. In repeated cycles, lower-frequency inversions usually terminate earlier but higher frequencies still deliver model updates. For an objective comparison of the joint inversion with the parallel difference method, the effects of insufficient iteration count are reduced by performing an extra cycle of baseline and monitor inversion (we call this approach "iterated" parallel difference (Maharramov and Biondi, 2014b)). The results of applying the iterated parallel difference to the twodatasets are shown in Figures 5(a) and 6(a). While cross-updating demonstrates certain robustness with regard to uncorrelated noise in the data and computational artifacts (note the significant quantitative improvement of reconstructed difference magnitudes in Figures 5(b) and 6(b)), the TV-regularized achieves further signifcant improvement by reducing oscilatory artifacts and honoring both smooth and blocky components of the model

difference.

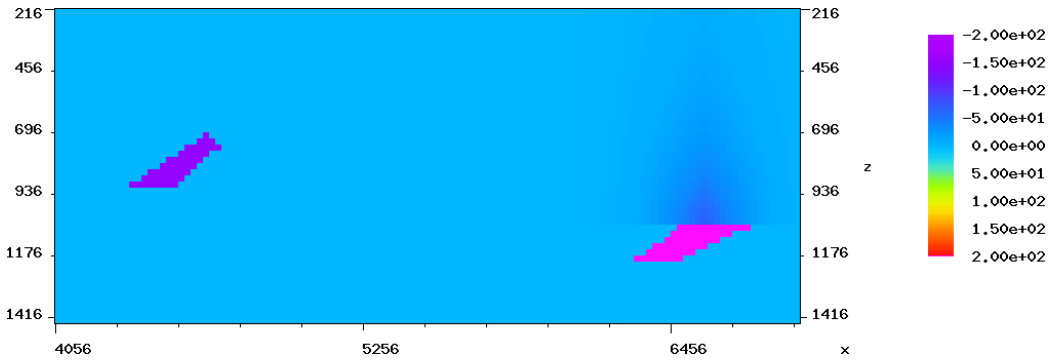


Figure 4: True velocity difference consists of a negative (−150 m/s) perturbation at about 4.5 km inline 800 m depth, and a positive (200 m/s) perturbation at 6.5 km inline, 1 km depth. [**CR**] musa1/. truediff

CONCLUSIONS

Our new TV-regularized simultaneous inversion technique is a more robust further development of our previous joint inversion method (Maharramov and Biondi, 2013, 2014b,a). Use of TV regularization in the simultaneous inversion allows recovery of production-induced changes without specifying variable weighting operator **W**, and penalizes unwanted model oscillations that may mask useful production-induced changes.

One potentially beneficial extension of this work is using TV regularization of individual models in (7,9). This approach may used for both time-lapse and standard FWI, and will be the subject future work.

ACKNOWLEDGMENTS

The authors would like to thank Stewart Levin for a number of useful discussions, and the Stanford Center for Computational Earth and Environmental Sciences for providing computing resources.

REFERENCES

Albertin, U., P. Sava, J. Etgen, and M. Maharramov, 2006, Adjoint wave-equation velocity analysis: 76th Annual International Meeting, SEG, Expanded Abstracts, 3345–3349.

Asnaashari, A., R. Brossier, S. Garambois, F. Audebert, P. Thore, and J. Virieux, 2012, Time-lapse imaging using regularized FWI: A robustness study: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi:10.1190/segam2012-0699.1, 1–5.

Aster, R., B. Borders, and C. Thurber, 2012, Parameter estimation and inverse problems: Elsevier.

Ayeni, G. and B. Biondi, 2012, Time-lapse seismic imaging by linearized joint inversion—A Valhall Field case study: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi:10.1190/segam2012-0903.1, 1–6.

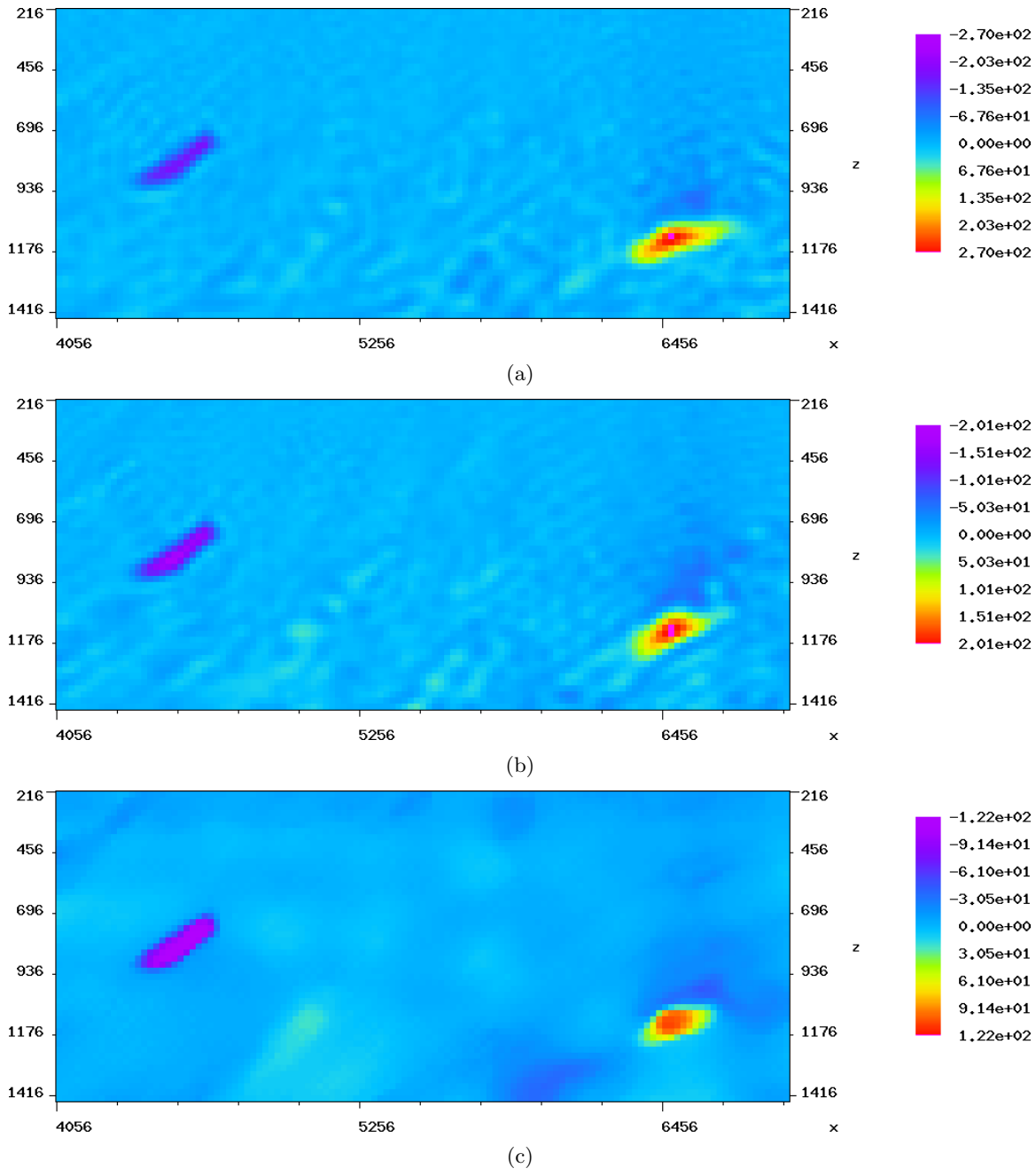


Figure 5: Model difference inverted from a clean synthetic with matching baseline and monitor acquisition geometries using (a) iterated parallel difference; (b) cross-updating; (c) TV-regularized simultaneous inversion. [**CR**] musa1/. clpd,clx,cltv

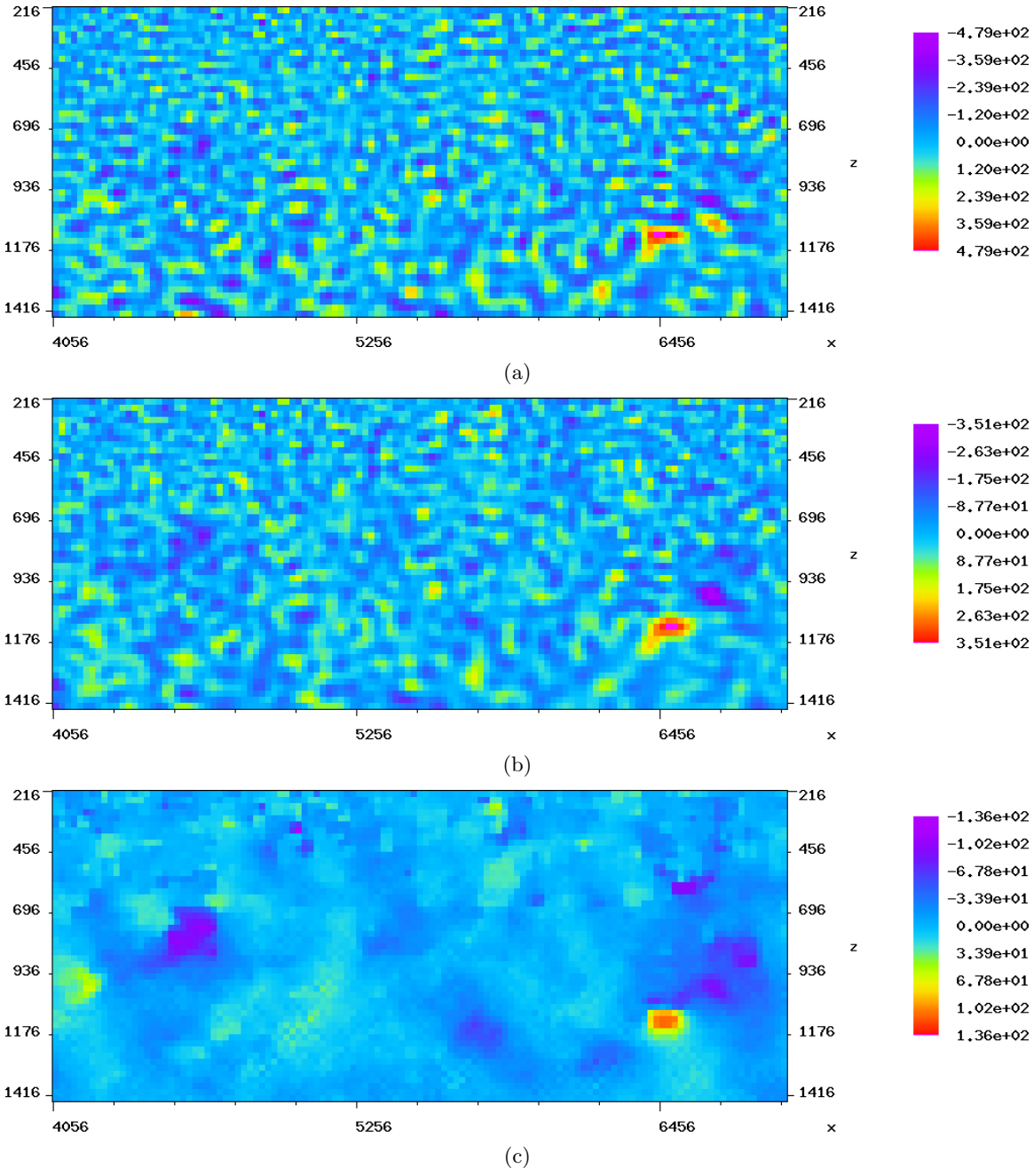


Figure 6: Model difference inverted from a 7 dB SNR synthetic with matching baseline and monitor acquisition geometries using (a) iterated parallel difference; (b) cross-updating; (c) regularized simultaneous inversion. [**CR**] musa1/. npd,nx,ntv

Barkved, O. and T. Kristiansen, 2005, Seismic time-lapse effects and stress changes: Examples from a compacting reservoir: The Leading Edge, **24**, 1244–1248.

Biondi, B., C. Deutsch, R. Gundes, D. Lumley, G. Mavko, T. Mukerji, J. Rickett, and M. Thiele, 1996, Reservoir monitoring: A multidisciplinary feasibility study: 66th Annual International Meeting, SEG, Expanded Abstracts, 1775–1778.

Denli, H. and L. Huang, 2009, Double-difference elastic waveform tomography in the time domain: 79th Annual International Meeting, SEG, Expanded Abstracts, 2302–2306.

Hatchell, P. and S. Bourne, 2005, Measuring reservoir compaction using time-lapse timeshifts: 75th Annual International Meeting, SEG, Expanded Abstracts, 2500–2503.

Maharramov, M. and U. Albertin, 2007, Localized image-difference wave equation tomography: 77th Annual International Meeting, SEG, Expanded Abstracts, 3009–3013.

Maharramov, M. and B. Biondi, 2013, Simultaneous time-lapse full waveform inversion: SEP Report, **150**, 63–70.

———, 2014a, Joint 4DFWI with model-difference regularization. SEG-AGU summer research workshop: SEG-AGU Summer Research Workshop. Advances in Active+Passive "Full Wavefield" Seismic Imaging: From Reservoirs to Plate Tectonics.

———, 2014b, Joint full-waveform inversion of time-lapse seismic data sets: SEP Report, **152**, 19–28.

Nocedal, J. and S. J. Wright, 2006, Numerical optimization: Springer.

Raknes, E., W. Weibull, and B. Arntsen, 2013, Time-lapse full waveform inversion: Synthetic and real data examples: 83rd Annual International Meeting, SEG, Expanded Abstracts, 944–948.

Routh, P., G. Palacharla, I. Chikichev, and S. Lazaratos, 2012, Full wavefield inversion of time-lapse data for improved imaging and reservoir characterization: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi:10.1190/segam2012-1043.1, 1–6.

Rudin, L. I., S. Osher, and E. Fatemi, 1992, Nonlinear total variation based noise removal algorithms: Phys. D, **60**, 259–268.

Sirgue, L., O. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. Kommedndal, 2010, Full waveform inversion: The next leap forward in imaging at Valhall: First Break, **28**, 65–70.

Sirgue, L. and R. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

Triebel, H., 2006, Theory of function spaces: Birkh auser.

Virieux, J. and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26.

Watanabe, T., S. Shimizu, E. Asakawa, and T. Matsuoka, 2004, Differential waveform tomography for time-lapse crosswell seismic data with application to gas hydrate production monitoring: 74th Annual International Meeting, SEG, Expanded Abstracts, 2323–2326.

Zheng, Y., P. Barton, and S. Singh, 2011, Strategies for elastic full waveform inversion of timelapse ocean bottom cable (OBC) seismic data: 81st Annual International Meeting, SEG, Expanded Abstracts, 4195–4200.

Advances in simultaneous salt boundary and background tomography updating

Taylor Dahlke, Biondo Biondi and Robert Clapp

ABSTRACT

Level set methods can provide a sharp interpretation of the salt body by defining the boundary as an isocontour of a higher dimensional implicit representation, and then evolving that surface to minimize the Full Waveform Inversion (FWI) objective function. Because the implicit surface update gradient is based on the tomographic update gradient, there is potential to utilize it to update the background velocity concurrently with the salt boundary. Using a shape optimization approach on synthetic examples, we can achieve reasonable convergence both in terms of the residual L2 norm, as well as the evolution of the salt boundary and background velocity towards the true model, demonstrating the feasibility of this approach. Various factors in processing the gradients and calculating step size influence this convergence, which we analyze and address. Ultimately, this method can be integrated into the processing work flow as a tool that provides improved building and refining of the velocity models used for imaging.

INTRODUCTION

Tomographic approaches to interpreting salt bodies can be less than effective, because the results tend to be too smooth to provide significantly accurate placement of salt boundaries. Manual and semi-automatic picking of salt boundaries is a common approach to interpreting the desired sharp delineations, but these methods can be time-consuming and tedious since expert input is necessary for either the actual picking, or the oversight and correction. Furthermore, once a model has been produced, it must be used to generate an image, and then be refined as necessary. A robust method for further automating the salt interpretation procedure would greatly alleviate this bottleneck.

Some previous approaches to interpreting salt boundaries use a shape optimization approach (Guo and de Hoop (2013), Lewis et al. (2012)). The boundaries of a salt body can be represented as the zero-isocontour of a higher dimensional surface (for example, a 2D boundary as a contour of a 3D surface). A gradient can be derived to evolve this shape / isosurface according to the FWI objective function. Unlike the smooth boundaries produced by tomographic approaches, the isocontour resulting from the shape optimization provides a sharp boundary, which is a more appropriate way to classify most salt-sediment interfaces. Guo and de Hoop (2013) utilize this approach using a frequency domain forward wave operator to evolve a salt boundary and velocity model. However, their approach alternates between updating the background velocity and salt body boundary, which effectively requires twice as many iterations as performing both updates concurrently.

The approach we take utilizes shape optimization with the use of time domain forward

wave-propagation, which allows us to take advantage of using a continuous range of frequencies (rather than discrete frequencies) in each iteration, allowing for sharper delineation of the boundary. Further, we take advantage of the fact that our boundary update gradient is based on the tomographic update gradient, and make updates to both concurrently after applying optimal scaling parameters. In theory, this method has the potential to be more efficient than an alternating update approach. In this paper we will discuss the fundamentals of the level set method and its key properties, followed by a demonstration of the concurrent boundary-tomography update method on a test model, as well as discussion on how we address the challenges inherent with concurrent updating.

THEORY

We begin with a brief overview of the level set method and how we apply the evolution scheme it utilizes. The full derivation for the shape optimization implementation can be found in Dahlke (2014).

Level set fundamentals

In our problem, we are trying to determine the boundary of a two dimensional body. Instead of using an algorithm that operates in this 2D plane directly, we use the level set algorithm which evolves a 3D implicit surface, ϕ. While our algorithm acts directly on this surface instead of the boundary, our solution for the 2D boundary is simply represented by a contour "slice" of this implicit surface where $\phi = 0$, as described in Osher and Sethian (1988) and Burger (2003). While it may seem counterintuitive to add extra dimensionality to our problem, by doing so we gain some advantages. These include the ability to merge and separate bodies as the level set evolution proceeds, as well as the ability to handle sharp corners and cusps in the lower-dimensional (2D) plane that the boundary exists on.

Based on this concept, we define a spatial domain $\Theta \subset \mathbb{R}^2$, a (salt) body $\Omega \subset \Theta$, and the salt body boundary Γ such that

$$\Omega = \{x \mid \phi(x, \tau) > 0\}, \qquad \Gamma = \{x \mid \phi(x, \tau) = 0\},$$

where τ indicates the axis along which the evolution steps progress ($\tau = 0$ is the initial iteration). As such, for a single step along τ, our salt body Ω evolves to $\Omega^{'}$. We define a point along the boundary curve to be

$$x_\Gamma = \{x \in \Gamma\}.$$

With this definition of the boundary points, the level set of ϕ that represents the salt body boundary can be described as

$$\phi(x_\Gamma, \tau) = 0.$$

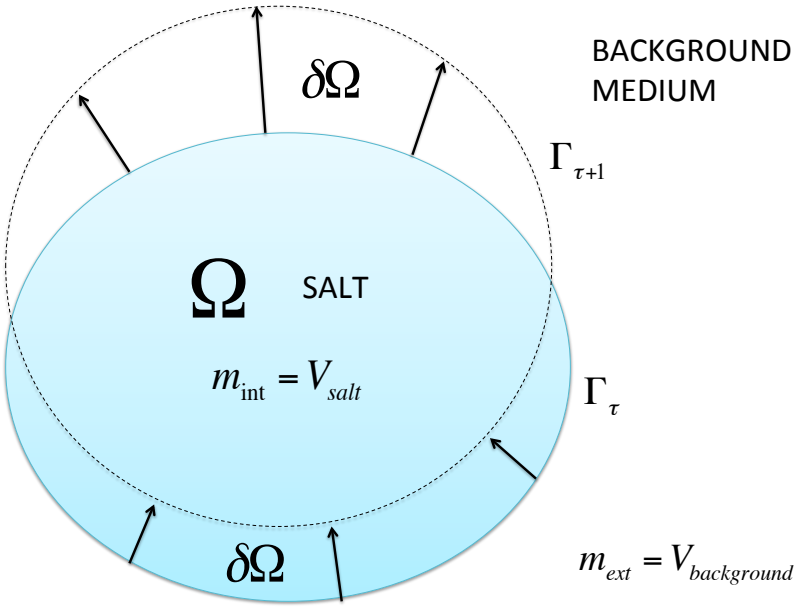


Figure 1: Diagram of domain partitioning. The full inclusive domain is Θ. Ω is the salt body. $\partial\Omega$ is the difference between the salt body domain in iteration τ and iteration $\tau+1$. Γ is the boundary of the salt body, with the subscript τ indicating the iteration. [**NR**] taylor/. domain-fig

By taking the derivative of this equation, the chain rule gives us

$$\frac{\partial \phi}{\partial \tau} + \frac{\partial \phi}{\partial x_\Gamma} \frac{\partial x_\Gamma}{\partial \tau} = 0. \tag{1}$$

This equation can be readily rewritten as

$$\frac{\partial \phi}{\partial \tau} + \vec{\nabla} \phi \cdot \vec{v}(x_\Gamma, \tau) = 0. \tag{2}$$

We can use $\vec{\nabla}\phi$ defined over all the full domain of x (rather than just x_Γ) since $\vec{\nabla}\phi \cdot \vec{v}(x_\Gamma, \tau)$ is a dot product, and only the terms where $x \in \Gamma$ will contribute to the overall dot product result. This "velocity" term in equation 2 can be defined as having both a "speed" and a normal vector component, $\vec{v}(x_\Gamma, \tau) = V(x_\Gamma, \tau)\vec{n}(x_\Gamma, \tau)$. In complete form there is also a tangential component, but we ignore this part since it doesn't contribute to a change in the surface ϕ.

We know the normal vector is defined as

$$\vec{n}(x_\Gamma, \tau) = \frac{\vec{\nabla}\phi(x_\Gamma, \tau)}{|\vec{\nabla}\phi(x_\Gamma, \tau)|} \ ,$$

which allows us to restate equation 2 in a more familiar representation

$$\frac{\partial \phi}{\partial \tau} = -V(x_\Gamma, \tau) \left| \nabla \phi \right| . \tag{3}$$

The scalar speed term $V(x_\Gamma, \tau)$ describes the magnitude of the variation of ϕ that is normal to the boundary Γ. It determines the evolution of the implicit surface, and ultimately the boundary implied by it. This term can be found to be

$$\frac{\partial \phi}{\partial \tau} = -\sum_s \int_0^T \int_{x \in \Gamma} (q_1^2 - q_2^2) h_s(x,t) \frac{\partial^2 u_s(x,t)}{\partial t^2} d\sigma dt \left| \nabla \phi \right| , \tag{4}$$

as described in the derivation provided in Dahlke (2014), which demonstrates how this formulation of the scalar speed term directs the evolution of the implicit surface such that it minimizes the FWI objective function. An important insight from this referenced derivation is that the scalar speed term contains the tomographic update gradient within it

$$-\sum_s \int_0^T \int_{x \in \Gamma} h_s(x,t) \frac{\partial^2 u_s(x,t)}{\partial t^2} d\sigma dt, \tag{5}$$

We can take advantage of it already being calculated and use it to make updates to the background velocity. In the following section we demonstrate this implementation approach.

APPLICATION ON SYNTHETIC EXAMPLES

We demonstrate the shape optimization algorithm on a 2D model, with the implicit surface evolved being a 3D surface. In this section we describe the algorithm used and the results that have come from its implementation.

Evolution algorithm

We begin with an initial background velocity, and a binary function as the initial implicit surface ϕ. Since we assume a constant salt velocity, we use both of these inputs to create a full initial-guess velocity model (m_o). Using this m_o, we forward model to get our d_{syn} which we use to get a residual. The residual is used to calculate both a tomographic and a boundary update gradient, as described in the derivation provided in Dahlke (2014). We then perform forward linearized operations on these gradients so that we can do a linear plane search (in residual space) for the scaling parameters α and β. Following this, we do a non-linear line search for a γ parameter that rescales α and β in a manner that minimizes the FWI objective function. We then apply an explict forward Euler scheme that updates the implicit surface (ϕ) and the background velocity V_{back}. This workflow is outlined in algorithm 1.

$$\phi^{j+1} = \phi^j + \gamma(\beta\frac{\partial\phi}{\partial j} + \mu G_{\text{reg}}), \tag{6}$$

$$V_{\text{back}}^{j+1} = V_{\text{back}}^j + \gamma\alpha\frac{\partial V_{\text{back}}}{\partial j}, \tag{7}$$

where β and α are the step sizes (for tomography and salt boundary respectively) and j is the current iteration point.

Scaling parameter optimization

As shown previously, the salt boundary update gradient is based on the adjoint of the linearized-Born operator, which is the tomographic update gradient. Since the gradient for both a tomographic and boundary update are calculated in each step regardless, we attempt to take advantage of this by finding scaling parameters to apply to these gradient updates such that we minimize the residual space objective function

$$\min_{\alpha,\beta} \left\| FG_{\text{tomo}}^T\alpha + FG_\phi^T\beta - (d_{\text{calc}} - d_{\text{obs}}) \right\|, \tag{8}$$

where F is the forward wave propagator, and G_{tomo}^T and G_ϕ^T are the update gradients for the background velocity and implicit surface ϕ respectively.

Minimizing this objective function gives us parameters that are scaled to the residual space, not the gradient space where they are actually applied. Since the adjoint operator that we use creates a scaling difference between the residual and gradient (data and model)

Algorithm 1 Shape optimization and tomographic update algorithm

Load observed data d_{obs}
Load initial implicit surface ϕ
Load background velocity V_{back}
for $i = 0$ to $numiter$ **do**
  Build full velocity model V_{full} from ϕ and V_{back}
  Forward model $\rightarrow d_{syn}$
  Calculate data space residuals ($r = d_{syn} - d_{obs}$)
  Perform direct arrival mute on residuals
  Back-propagate residuals (RTM imaging) $r \rightarrow \partial V_{full}/\partial i$
  Calculate $\nabla\phi$
  Calculate $\partial\phi/\partial i$
  Mute and smooth $\partial V_{full}/\partial i$
  Forward model $\partial\phi/\partial i \rightarrow d_{bound}$
  Forward model $\partial V_{full}/\partial i \rightarrow d_{tomo}$
  Peform linear plane search for α and β
  Perform non-linear plane search for γ
  Update boundary: $\phi^{i+1} = \phi^i + \gamma\beta\frac{\partial\phi}{\partial i}$
  Update background velocity: $V_{back}^{i+1} = V_{back}^i + \gamma\alpha\frac{\partial V_{full}}{\partial i}$
end for
Output final velocity model

spaces, we must rescale α and β once they are found so that they can be effectively applied to the gradients.

The approach we use is to rescale α and β according to a γ parameter which is found using a non-linear line search (which is constrained by the conditions for stability). This technique is much cheaper than performing a full non-linear plane search for α and β, but still allows for a choice of parameters based on the FWI objective function. We choose this approach, utilizing equation 9 for a line search for γ,

$$\min_{\gamma} \|F(m(\gamma) - d_{\text{obs}})\| . \tag{9}$$

Tomographic gradient masking and smoothing

The separation of the tomographic information from the reflectivity information is desired so that the tomogaphic updates more quickly lead to convergence of the true solution. This seperation can be better achieved prior to this search for the scaling parameters by masking out the tomographic update gradient in areas where the update has no influence on changing the next iteration of the velocity model. For example, in this work we assume a constant velocity throughout the salt bodies we model. Because of this, we dont apply the tomographic gradient update in the regions where salt exists. If we calculate G_{tomo}^T without first masking out G_{tomo} in areas overlapped by salt regions, then we introduce bias into the objective function (equation 8), since it will optimize for an update that will not be entirely applied.

We further assume that the salt boundary change will not undergo significant shifts. With this in mind, we apply the masking based on the salt body delineation that was created from the most recent (previous) iteration. Another approach would be to dynamically update the salt boundary based on the scaling parameter β, as β is being solved for. While theoretically producing a more accurate update, this method is also far more expensive, since numerous applications of the forward linearized-Born operator are necessary. For this reason we make the approximation of masking based on the previous iteration of the salt boundary.

When the masking is performed and the salt boundary shrinks, an area of the background velocity is exposed which contains a sharp boundary between the newly "exposed" region and the region that was previously exposed and updated. This can create false (albeit weak) reflectors around the edge of the salt, causing errors as the evolution of the salt boundary continues. For this reason, immediately after masking is performed on the tomographic gradient, a smoothing operator is applied to remove sharp discontinuities in the velocity update along this boundary. Because the tomographic update information tends to be lower frequency than reflection information, this step also helps separate reflection and tomographic information by acting as a low-pass filter.

Stability

As the implicit surface is evolved, it is important to maintain stability of the evolution. One relevant aspect of maintaining stability is keeping the implicit surface update step size (β) small enough to satisfy the Courant-Friedrich's-Levy (CFL) condition, which is stated by Chaudhury and Ramakrishnan (2007) (when applied to level set evolution) as being

$$G_{\max} \cdot \gamma\beta \leq \min(h_x, h_y), \tag{10}$$

where h_x and h_y are the grid spacing in the x and y directions, and $G_{\max}$ is the maximum value of the update gradient. While later we describe how a plane search is used to determine the scaling parameters α and β, our algorithm adjusts these scaling parameters (while maintaining their ratio) in such a manner that satisfies the constraint in equation 10.

An additional measure taken to ensure the stablity of evolution is the use of a regularization term that is scaled and added to the boundary gradient before each update is applied. In this case, a distance regularization term was used. This term drives the spatial gradient of the implicit surface towards either one or zero ($\nabla\phi = \{1, 0\}$).

When irregularities begin to occur in the implicit surface during level set evolution, numerical errors start to occur which can lead to instability. By regularizing the gradient of the implicit surface as it evolves, we minimize irregularities and are able to continue evolution without having to reinitialize a signed-distance function to the salt boundary contour. An excellent reference on this type of regularization is Li et al. (2010).

Results

We apply our algorithm on a simple velocity model, using an acquisition geometry of 32 shots spaced 110 [m] apart, and 63 receivers spaced 50 [m] apart. In the example shown

in Fig. 2, the initial and true background velocity models differed by up to 100 [m/s] (see Fig. 5). A bottom reflector and positive velocity gradient provides better illumination along the bottom and flanks of the circular salt body, which has a velocity of 4500 [m/s]. A stencil radius of five was used for smoothing of the tomographic update gradient prior to its application.

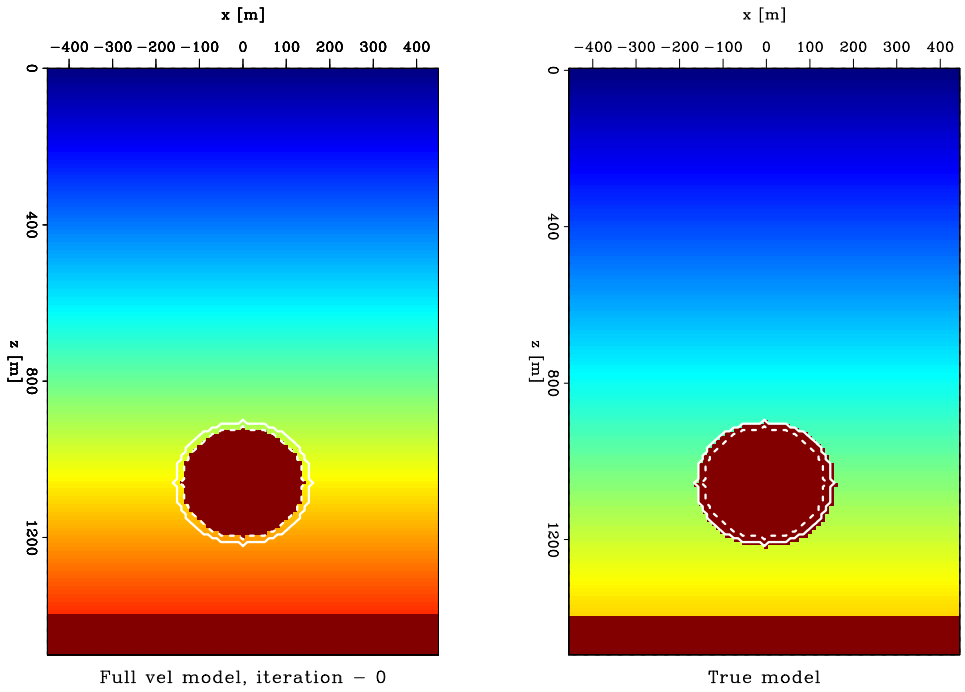


Figure 2: Initial velocity model (left) and true velocity model (right). True model boundary indicated (solid line); Initial boundary guess (dashed line). [**ER**] taylor/. SBS-frames-149-zoom

DISCUSSION

One thing to note when observing the trend in Fig. 4 is that the objective function does not decrease monotonically, and in some cases increases slightly. The non-linear line search for γ finds a minimum to the objective function by doing a quadratic interpolation between the three points surrounding the working minimum. This approximation is done for efficiency since the non-linear calculation for each point is relatively expensive, so a relatively low density of sampling along the γ axis is done. Because of the error inherent in this approximation, the algorithm may find a step size that increases the objective function slightly rather than choose a step size equal to zero.

Another observation is that the objective function plateaus at a solution that has a very low residual error, but is not the correct answer, indicating the algorithm has converged on a local minima (Fig. 4). A number of characteristics of the algorithm play into this. One is the masking operation that is performed on the tomographic update gradient. The RTM imaging places energy for the updates based on both reflection and tomographic discrepancies in the model. This mixing of reflection and tomography information is inherent

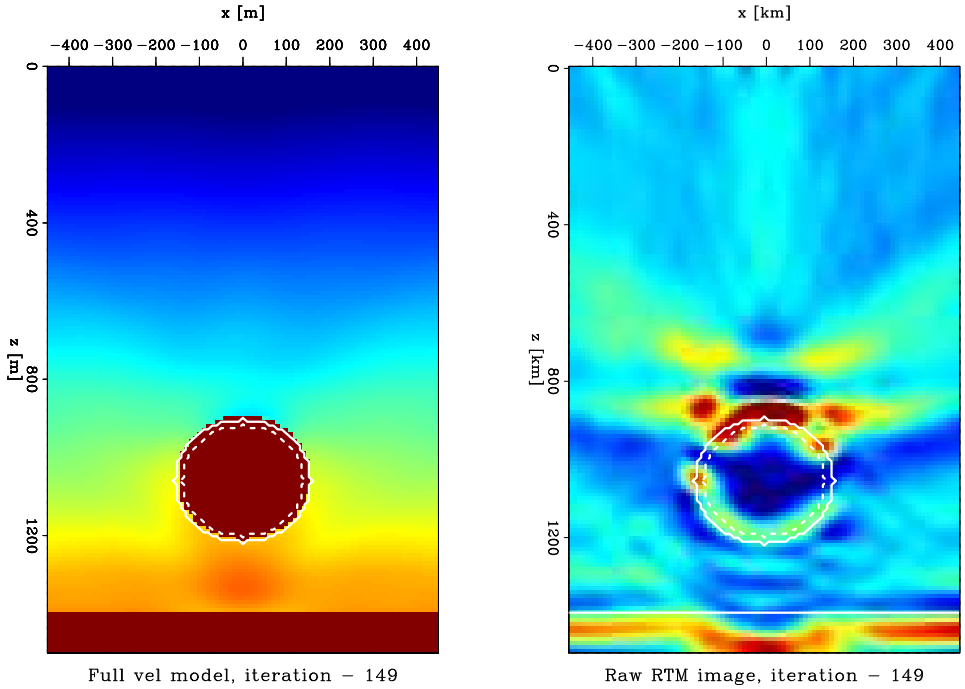


Figure 3: Final velocity model (left) and raw RTM image (right) after 149 iterations. [**ER**] taylor/. figure5

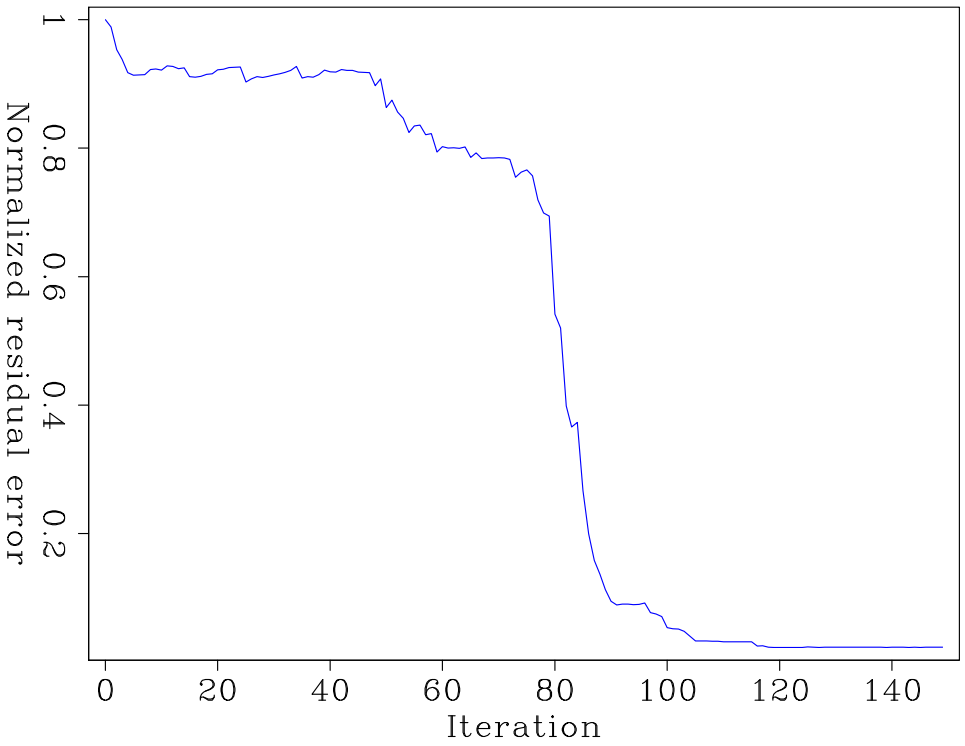


Figure 4: Objective function for evolution shown in Fig. 2 [**ER**] taylor/. obj-val

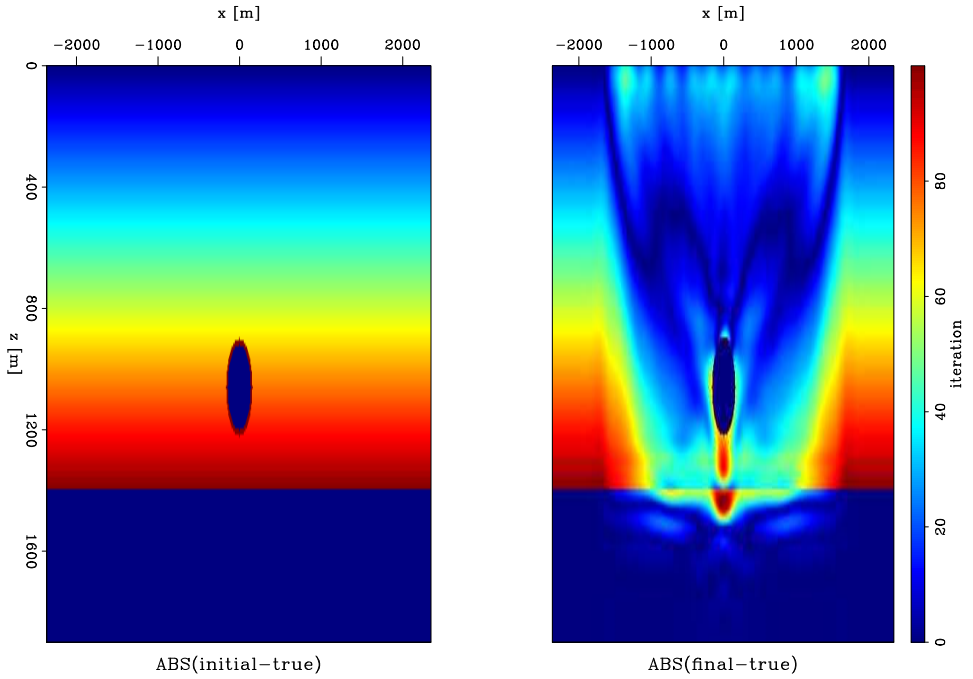


Figure 5: Change in the velocity difference for the model shown in Fig. 2 [**ER**] taylor/. vel-diff

to the concurrent updating approach. Even when the salt boundaries come quite close to the true position, the imaging may place energy inside the salt body region (See Fig. 3), which is never translated into a model update due to the assumption of constant velocity salt. In a model such as the one demonstrated on, the paucity of non-salt reflectors exasperates this problem and leads to convergence at a local minima solution, since some of the strongest updating energy is masked over during updating.

A further challenge demonstrated in this model is the lack of updating directly along the bottom reflector (See Fig. 3). One explanation for this is the band-limited nature of the experiment. Because the reflector acts as a step function in velocity space, in order to perfectly resolve it all frequencies must be available. Of course, our experiment uses a source wavelet centered at 15 Hz, which means our imaging will not be able to resolve this boundary to satisfaction unless more high frequencies are included.

The density of acquisition and geometry of the array also influence the number and strength of RTM imaging artifacts, which also influences the rate and quality of convergence. While a higher density of acquisition will typically lead to fewer artifacts and smoother tomographic updates, the smoothing of the tomographic update gradient also simulates that effect and is more computationally efficient. Figures 6 and 7 show that using a longer stencil length allows for quicker convergence, and also a better match to the true model. However, this approach has its limitations, and can reduce the ability of the algorithm to resolve sharper tomographic anomalies, exasperating the problems created by the band limited nature of our experiment.

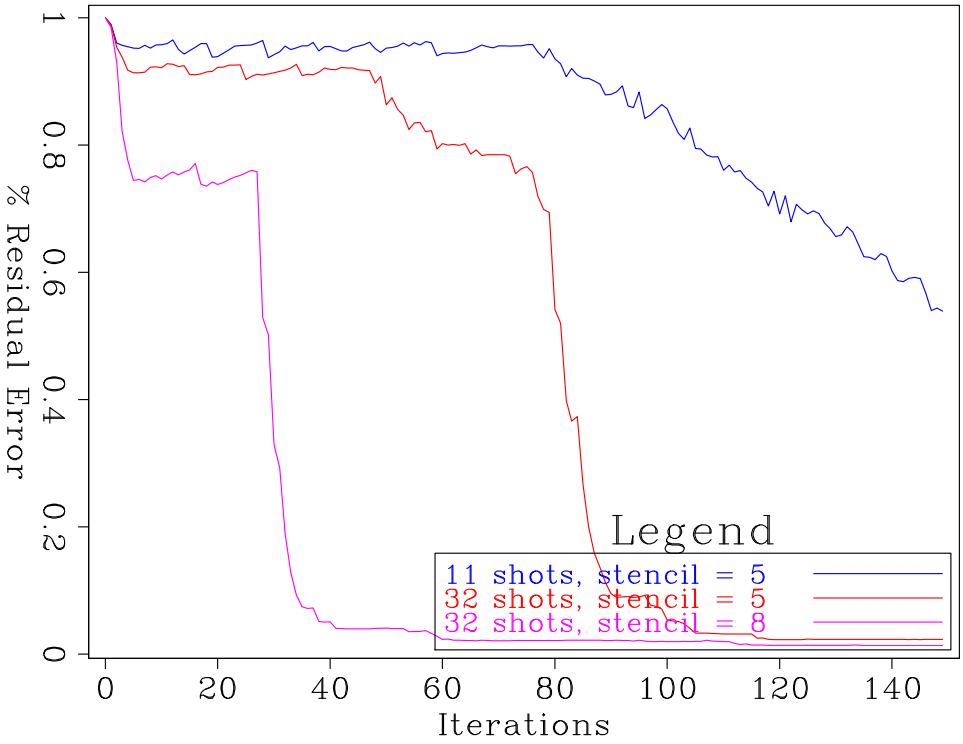


Figure 6: Different rates of convergence using the same model from Fig. 2 [**ER**] taylor/. smooth-compare

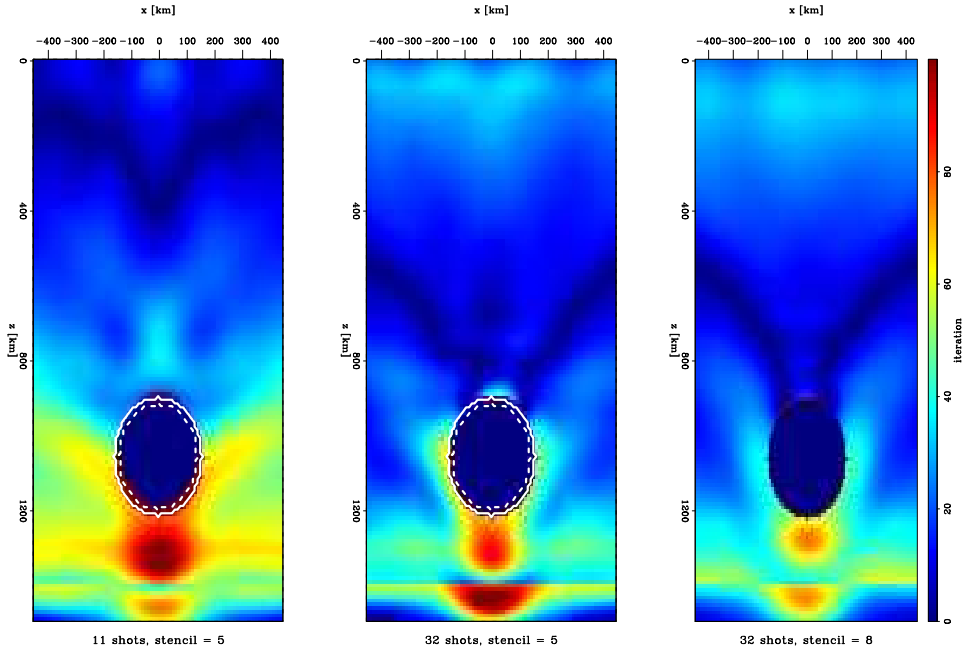


Figure 7: The absolute value of the difference between the true model and the final solution $\|V_{true} - V_{final}\|$ for varying smoothing stencil lengths and array densities. [**ER**] taylor/. smooth-compare2

CONCLUSION

In this work we described the derivation of the level set method as applied to the minimization of the FWI objective function. We demonstrated the application of this evolution algorithm and its incorporation with a background velocity tomographic update on a simple model. We consider the limitations of this approach in regards to numerical stability, as well as the assumptions of linearity that we use to find our scaling parameters, and the challenges regarding the separation of reflection information from the tomographic updating.

ACKNOWLEDGEMENTS

I would like to thank Adam Halpert, Ali Almomin, Musa Maharramov, Sjoerd de Ridder, and Stew Levin for their insight into this work. Further, I thank the SEP sponsors for their generous financial support of this research.

REFERENCES

Burger, M., 2003, A framework for the construction of level set methods for shape optimization and reconstruction: Interfaces and Free boundaries, **5**, 301–330.

Chaudhury, K. N. and K. R. Ramakrishnan, 2007, Stability and convergence of the level set method in computer vision: Pattern Recogn. Lett., **28**, 884–893.

Dahlke, Biondi, C., 2014, Shape optimization using the fwi objective function for salt body segmentation: SEP-152, 29–44.

Guo, Z. and M. de Hoop, 2013, Shape optimization and level set method in full waveform inversion with 3d body reconstruction: SEG Technical Program Expanded Abstracts, 1079–1083.

Lewis, W., B. Starr, D. Vigh, et al., 2012, A level set approach to salt geometry inversion in full-waveform inversion: Presented at the 2012 SEG Annual Meeting.

Li, C., C. Xu, C. Gui, and M. Fox, 2010, Distance regularized level set evolution and its application to image segmentation: Image Processing, IEEE Transactions on, **19**, 3243–3254.

Osher, S. and J. A. Sethian, 1988, Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations: Journal of computational physics, **79**, 12–49.

SEP PHONE DIRECTORY

Name	Phone	Login Name
Almomin, Ali	723-0463	ali
Alves, Gustavo	723-9282	gcalves
Barak, Ohad	723-9282	ohad
Barnier, Guillaume	723-6006	gbarnier
Biondi, Biondo	723-1319	biondo
Biondi, Ettore	725-1625	ettore
Blatter, Daniel	724-9339	dblatter
Cabrales, Alejandro	723-3187	cabrales
Chang, Jason	724-4322	jason
Claerbout, Jon	723-3717	jon
Clapp, Bob	725-1334	bob
Dahlke, Taylor	724-4322	taylor
Lau, Diane	723-1703	diane
Le, Huy	723-1250	huyle
Leader, Chris	723-0463	chrisl
Levin, Stewart	726-1959	stew
Maharramov, Musa	723-1250	musa
Martin, Eileen		ermartin
Ruan, Kaixi		kaixi
Shen, Yi	723-1250	yishen
Wong, Mandy	723-9282	mandyman
Zhang, Yang	723-6006	yang

Dept fax number: (650) 725-7344

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

Sponsors who have provided us with their domain names are not prompted for a password when they access from work. If you are a sponsor, and would like to access our restricted area away from work, visit our website and attempt to download the material. You will then fill out a form, and we will send the username/password to your e-mail address at a sponsor company.

STEERING COMMITTEE MEMBERS, 2014-2015

Name	Company	Tel #	E-Mail
Raymond Abma	BP	(281)366-4604	abmar1@bp.com
Francois Audebert (Co-chair, 2nd year)	TOTAL	–	francois.audebert@total.com
Biondo Biondi	SEP	(650)723-1319	biondo@sep.stanford.edu
Jon Claerbout	SEP	(650)723-3717	jon@sep.stanford.edu
Thomas Dickens	ExxonMobil	(713)431-6011	tom.a.dickens@exxonmobil.com
Faqi Liu	Hess	–	fliu@hess.com
Yi Luo	Saudi Aramco	–	yi.luo@aramco.com
Alejandro Valenciano	PGS	–	alejandro.valenciano@pgs.com

Research Personnel

Ali Almomin graduated from Texas A&M University in 2007 with a BS in Geophysics. Then, he joined Saudi Aramco and worked in several exploration and research departments with a focus on 3D seismic processing and near surface imaging. He joined Stanford Exploration Project in 2009 to pursue a PhD in Geophysics and is currently working on seismic tomography. He is a member of SEG, EAGE, and SPE.

Gustavo Catao Alves graduated from the Federal University of Minas Gerais, Brazil, in 2004 with a BS in Physics. In 2006, he got a MS in Solid State Physics from the same Institution and was awarded Best Student Presenter on the 3rd Conference of the Brazilian Carbon Association. That same year he joined Petrobras, where he worked at the Company's research center for the next 7 years, on projects focused on seismic illumination studies, Reverse Time Migration and Full Waveform Inversion. In 2008, he received an Honorable Mention for Best Poster Paper at SEG's Annual Meeting, with the work entitled "Target oriented illumination analysis using wave equation". He joined Stanford Exploration Project in 2013 to pursue a PhD in Geophysics and is currently working on amplitude studies in long offset data. He is a member of SBGf, SEG, EAGE and AAPG.

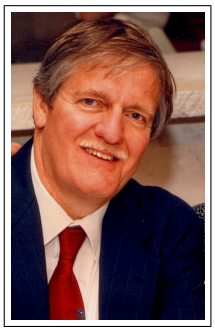
Ohad Barak received a B.Sc. (2006) and an M.Sc. (2009) in Geophysics from Tel-Aviv University. In 2008 he joined the R&D team of Paradigm Geophysical, maintaining and developing the production codes. He joined SEP in 2009 and is currently pursuing a Ph.D. in geophysics at Stanford University, and a longer biography. His current research is about wave-mode separation of multi-component seismic data that includes both displacement and rotational data.

Guillaume Barnier is from Nice, France. He graduated from Telecom Paristech in 2007 with a MSc in telecommunications and signal processing. After working as a fixed income trader for JP Morgan (London) from 2007 to 2010, he decided to change career path by joining the geophysics MSc program in Colorado School of Mines (2011), where he focused his research on seismoelectric coupling in poro elastic media. In September 2013, he joined the Stanford Exploration Project to pursue his Ph.D. So far, he has been working on quantifying structural uncertainty in seimsic images due to error in velocity estimation.

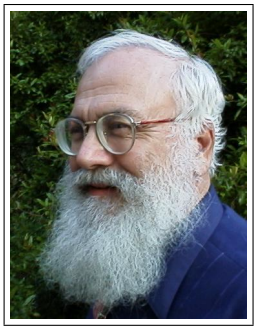
Biondo L. Biondi is professor of Geophysics at Stanford University. Biondo graduated from Politecnico di Milano in 1984 and received an M.S. (1988) and a Ph.D. (1990) in geophysics from Stanford. He is co-director of the Stanford Exploration Project and of the Stanford Center for Computational Earth and Environmental Science. In 2004 the Society of Exploration Geophysicists (SEG) has honored Biondo with the Reginald Fessenden Award. Biondo recently published a book, 3-D Seismic Imaging, that is the first text book to introduce the theory of seismic imaging from the 3-D perspective. The book is published by SEG in the Investigations in Geophysics series. During 2007 gave a one-day short course in 28 cities around the world as the SEG/EAGE Distinguished Short Course Instructor (DISC) . He is a member of AGU, EAGE, SEG and SIAM.

Ettore Biondi received a B.S. (2010) in geology from the University of Genoa, an M.S. (2012) in geophysics from the University of Pisa, and a diploma (2013) in computational chemistry from the Scuola Normale Superiore of Pisa. He spent almost two years as a research fellow at the University of Milan within the geophysics section of the Earth Science Department. In the summer of 2014, he joined the SEP and started his Ph.D. at the Stanford University.

Robert Brune is a Geophysical Consultant who has become an advocate for rotational seismic applications in E&P since the mid-2000s, inspired by the early work in earthquake seismology. He is a believer in the revolutionary potential of full multi-component utilization of seismic wavefields, and has done work in testing and advocating the use of rotational measurements in spatial sampling, mode selection, frac & injection monitoring, S/N processing, etc. Bob has worked at GSI, the USGS (at Menlo), Sohio/BP, and TGS, and consulting for service and operating companies, both in technology and operations. He has worked in both E&P, including as an Exploration Mgr., and Regional Field Development Mgr. He has worked in technology as a Geophysical Technical Services and R&D Mgr. In the services industry he was President-Offshore, and Chief Geophysicist at TGS. Robert has a B.S. in Geology from U. of Missouri at Rolla where he graduated at 19, an M.S. in Geophysics from Stanford where he was present at the startup of SEP, an M.S. in Computer Systems from U. of Denver, and extensive coursework in Petroleum Eng., math, chemistry, and engineering.



Jon F. Claerbout (M.I.T., B.S. physics, 1960; M.S. 1963; Ph.D. geophysics, 1967), professor at Stanford University, 1967. Emeritus 2008. 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.

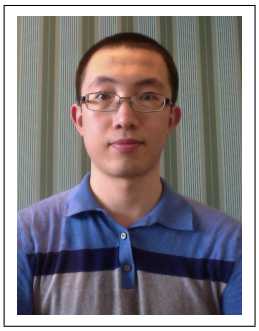
Taylor Dahlke is a third year student with SEP. He received his B.S. in civil engineering from the University of California, Berkeley in 2012, and joined SEP in July 2012. Currently, he is working towards a Ph.D. in geophysics with his research focused on applying levelset methodologies to perfom salt body image segmentation. Taylor is a student member of SPE and Vice President of the Stanford SEG Student Chapter.

Carlo Fortini is from Caravaggio, Italy. He graduated from Politecnico di Milano with a bachelor degree in Telecommunication engineering in 2008. In 2011 he received his MSc from Politecnico di Milano with a thesis in the field of geophysics and he then enrolled in the PhD program at the same school. The main focus of his work is the processing and imaging of the multiple reflections. His work has been done in collaboration with the italian oil company eni. During the PhD studies he also worked as an internship student at WesternGeco (London) and he spent six months as a visiting student researcher at the Stanford Exploration Project. He is going to graduate in November 2014. Carlo is a member of SEG and EAGE.

Antoine Guitton received a M.Sc. in geophysics from Université de Strasbourg in 1996. He also received a M.Sc. and Ph.D in geophysics from Stanford University in 1998 and 2005. He is a recipient of the EAGE Arie van Weelden Award in 2004 and SEG J. Clarence Karcher Award in 2007. He received the Best Paper Presented by a Student at the Annual Meeting Award in 1999 for his work on robust norms. Antoine is currently working for Geoimaging Solutions Inc. as a senior research geophysicist working on imaging and velocity estimation problems. He was a consulting professor of geophysics at Stanford University from August 2005 to August 2014. He is a member of the EAGE, SEG and AGU.

Chris Leader graduated from Oxford Univeristy in 2008 with a BA in Physics (with concentration on Astrophysics and Condensed Matter physics) and then from Imperial College London in 2009 with an MSc in Petroleum Geophysics (Distinction). He is currently a first year student in the Stanford Exploration Project on the PhD program working on Fourier methods of regularisation. Work experience involves 3D seismic processing for a Rio Tinto acquired dataset over summer 2009. He is a member of SEG, EAGE, PESGB and IOP.

Stewart A. Levin was acting director of the Stanford Exploration Project during Jon Claerbout's 1993-4 sabbatical year. After a distinguished career in industry at Mobil and Halliburton, he has returned to Stanford as a consulting professor in the Department of Geophysics.



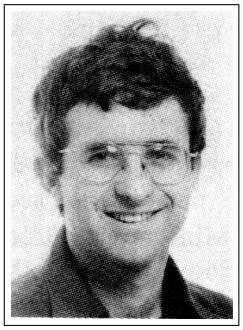
Yinbin Ma graduated from University of Science and Technology of China in 2009 with a B.S. in physics. In 2013, he obtained a M.S. in Condensed Matter Theory at Rice University. Yinbin moved to Stanford in 2013 and is currently working towards a Ph.D. in the Institute for Computational and Mathematical Engineering. He started doing research at SEP from June 2014. He is a student member of the SEG.

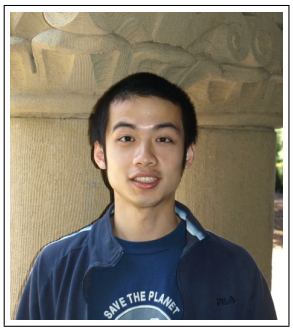
Musa Maharramov received his diploma c.l. in Mathematics from Baku State University in 1993. In 1995, he received a Ph.D. in Mathematics (Differential Equations) from the Azerbaijan Institute of Mathematics for his thesis "Asymptotic Solutions of Quasilinear Parabolic Equations". In 1995-2001, he worked as an IT& Comms consultant, later IT manager, at Fluor Corporation in Baku. In 2001, he joined BP Azerbaijan, and in 2006 he started his work at BP High Performance Computing as a Computational Scientist. In that role he worked with the BP Imaging Team on mathematical, algorithmic and computational aspects of seismic migration and inversion. Musa joined SEP in the fall of 2011 and is pursuing his second Ph.D., in Geophysics. He is currently working on regularization of seismic inversion problems through the application of Geological/Geomechanical constraints and advanced optimisation techniques. Musa is a member of the SEG, EAGE and SIAM.

Paul Milligan graduated from the University of California at Berkeley with both M.Sc. and Ph.D. degrees in Engineering Geoscience in 1999. He has worked extensively in the petroleum industry specializing in seismic data acquisition and processing, including marine streamer seismic, ocean bottom seismic, VSP, and crosswell seismic, while working for the following companies: GSI, Sonics Exploration, Geco-Prakla, Baker Hughes, Paulsson Geophysical services, Seabird Exploration, and currently Seabed Geosolutions. He also has experience working further downstream on velocity inversion and wave equation migration imaging while working for 4th Wave Imaging and 3DGeo. He is currently working offshore for Seabed Geosolutions, processing ocean bottom seismic data, and developing quality control methods and software for the multi-component data.

Sjoerd de Ridder hails from the Netherlands and received a BSc (2004) in Earth Sciences from Utrecht University and an MSc (2007) in Applied Earth Sciences from Delft University of Technology. In the fall of 2007 he started a PhD program at Stanford University to work on surface-wave seismic-interferometry for reservoir scale imaging in the Stanford Exploration Project. He received the Jon F. Claerbout - Chevron Fellowship in 2009-2010. During his PhD program he spent summers at China University of Petroleum in Beijing (2010), BP America (2011), Chevron (2012) and BP Norway (2013). He graduated in March of 2014 and started as an Associate Research Professor at the University of Science and Technology of China (USTC) in Hefei, Anhui. Supported by Chevrons Center of Research Excellence (CORE) at Stanford, Sjoerd escaped the summer heat of Hefei and spend July and August in California. In his free time he likes to run, bike and hike while traveling and discussing philosophy, politics and history. In recent years he cycled over 3000 miles exploring China.

Shuki Ronen is currently (2013) the chief geophysicist of Seabed Geosolutions, and a consulting faculty at SEP. The photo dates back to when he was a student at Stanford, 1985.





Xukai Shen graduated from Tsinghua University in July 2007 with a B.E. in Electrical Engineering. He joined SEP in the fall of 2007, and is currently working toward a Ph.D. in Geophysics. He is a student member of the SEG.

Mandy Wong graduated in 2004 with a B.Sc. in Physics and Mathematics from the University of British Columbia (UBC) in Vancouver, Canada. In 2006, she obtained a M.Sc. degree in Condensed Matter Thoery at UBC. Afterward, Mandy worked for a geophysical consulting company, SJ Geophysics, based in Vancouver, Canada. Mandy joined SEP in 2008, and is working towards a Ph.D. in Geophysics. Her main research interest is imaging with multiples.

SEP ARTICLES PUBLISHED OR IN PRESS

Almomin, A., and Biondi, B., 2012, Tomographic full waveform inversion: Practical and computationally feasible approach. SEG Technical Program Expanded Abstracts, 1–5.

Almomin, A., and Biondi, B., 2013, Tomographic full waveform inversion (TFWI) by successive linearizations and scale separations. SEG Technical Program Expanded Abstracts, 1–5.

Almomin, A., and Biondi, B., 2014, Preconditioned tomographic full waveform inversion by wavelength continuation. SEG Technical Program Expanded Abstracts, 1–5.

Barak, O., Jaiswal, P., de Ridder, S., Giles, J., Brune, R. and Ronen, S., 2014, Six-component seismic land data acquired with geophones and rotation sensors: wave-mode separation using 6C SVD: SEG Technical Program Expanded Abstracts (Accepted).

Barak, O., Herkenhoff, F., Dash, R., Jaiswal, P., de Ridder, S., Giles, J., Brune, R. and Ronen, S., 2014, Six-component seismic land data acquired with geophones and rotation sensors: wave-mode selectivity by application of multicomponent polarization filtering: The Leading Edge (Accepted).

Chang, J., de Ridder, S., and Biondi, B., 2013, Power spectral densities and ambient noise cross-correlations at Long Beach: SEG Technical Program Expanded Abstracts **32**, 2196–2200.

Chang, J., Nakata, N., Clapp, R.G., de Ridder, S., and Biondi, B., 2014, High-frequency surface and body waves from ambient noise cross-correlations at Long Beach, CA: SEG Technical Program Expanded Abstracts **33**(Accepted).

Dahlke, T., Beroza, G., Chang, J., and de Ridder, S., 2014, Stochastic variability of velocity estimates using eikonal tomography on the Long Beach data set: SEG Technical Program Expanded Abstracts **33** (Accepted).

de Ridder, S.A.L., B.L. Biondi and R.G. Clapp, 2014, Time-lapse ambient-seismic-noise tomography at Valhall, SEG Technical Program Expanded Abstracts **33** (Accepted).

de Ridder, S.A.L., B.L. Biondi and R.G. Clapp, 2014, Time-lapse seismic noise correlation tomography at Valhall, Geophysical Research Letters, Vol. **41**, doi:10.1002/2014GL061156, 61166122.

de Ridder, S. and B. Biondi, 2014, Continuous reservoir-scale passive monitoring using microseism noise: SEG/AGU Summer Research Workshop, Vancouver, Canada, 22-24 June.

de Ridder, S., B. Biondi and D. Nichols, 2014, Passive seismic interferometry and anisotropic eikonal phase-velocity tomography with examples from Ekofisk field: 76th EAGE Conference & Exhibition 2014, Amsterdam, NL, 16-19 June 2014, Th E102 13.

de Ridder, S.A.L. and B.L. Biondi, 2013, Daily reservoir-scale subsurface monitoring using ambient seismic noise, Geophysical Research Letters, Vol. **40**, doi:10.1002/grl.50594, 29692974.

Guitton, A., Ayeni, G., and Esteban, D.A., 2012, Constrained full-waveform inversion by model reparameterization: Geophysics, 77, No. 2, R117–R127.

Guitton, A., and Esteban, D.A., 2012, Attenuating crosstalk noise with simultaneous source full waveform inversion: Geophysical Prospecting (In Press).

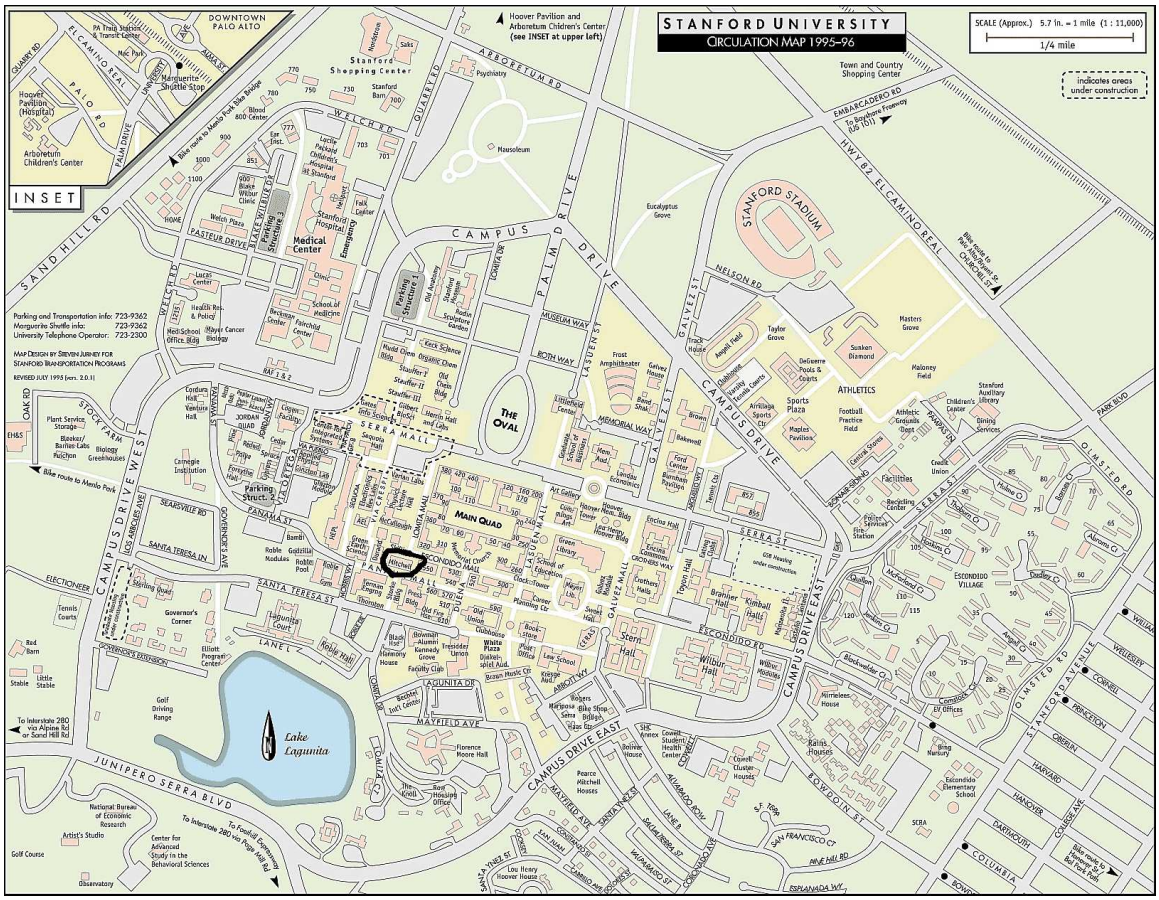
Guitton, A., 2012, Blocky regularization schemes for full waveform inversion: Geophysical Prospecting (In Press).

Halpert, A., 2012, Edge-preserving smoothing for segmentation of seismic images: SEG Technical Program Expanded Abstracts **31**, 1–5.

Halpert, A., Clapp, R.G., and B. Biondo, 2014, Salt delineation via interpreter-guided

seismic image segmentation, Interpretation: **2**, T79–T88.
Arevalo, Humberto and Stewart A. Levin, 2014, Well and Seismic matching with ArcGIS and ProMAX via KML: Esri International User Conference Paper 901, San Diego, 15 July.
Leader, C., and R. Clapp, 2012, Least squares reverse time migration on GPUs - balacing IO and computation: 74th Conference & Technical Exbition, EAGE, Extended Abstracts (Accepted).
Leader, C., and A. Almomin, 2012, How incoherent can we be? Phase encoded linearised inversion with random boundaries: SEG Technical Program Expanded Abstract **31** (Submitted).
Levin, Stewart A. and Fritz Foss, 2014, Downward continuation of Mars SHARAD data: SEG Technical Program Expanded Abstract **33** (Submitted).
Li, Y., Y. Zhang, and J. Claerbout, 2012, Hyperbolic estimation of sparse models from erratic data: Geophysics **77**, 1–9.
Li, Y., P. Shen, and C. Perkins, 2012, VTI migration velocity analysis using RTM: SEG Technical Program Expanded Abstract **31**.
Li, Y., Image-guided WEMVA for azimuthal anisotropy: SEG Technical Program Expanded Abstract **32**.
R.E. Plessix and Y. Li, 2013, Waveform acoustic impedance inversion with spectral shaping: Geophysical Journal International **195**(1), 301–314, 2013.
Li, Y., B. Biondi, D. Nichols, and R. Clapp, 2014, Wave equation migration velocity analysis for VTI models: Accepted for publication in Geophysics, 2014.
Li, Y., B. Biondi, D. Nichols, and R. Clapp, 2014, Wave equation migration velocity analysis for VTI models: Accepted for publication in Geophysics, 2014.
Li, Y., M. Wong, and R. Clapp, 2014, Equivalent accuracy by fractional cost: overcoming temporal dispersion: Submitted to SEG 2014.
Li, Y., D. Nichols, and G. Mavko, 2014, Stochastic rock physics modeling for seismic anisotropy: Submitted to SEG 2014.
Li, Y., B. Biondi, R. Clapp, and D. Nichols, 2014, Rock physics constrained anisotropic wave-equation migration velocity analysis: Submitted to SEG 2014.
Martin, E., 2014, Improving memory bandwidth usage in forward modeling with compression: SEG Technical Program Expanded Abstracts (Submitted).
Shen, Y., B. Biondi, R. Clapp, and D. Nichols, 2013, Wave-equation migration Q analysis (WEMQA): EAGE Workshop on Seismic Attenuation Extended Abstract 2013.
Shen, Y., B. Biondi, R. Clapp, and D. Nichols, 2014, Wave-equation migration Q analysis (WEMQA): SEG Technical Program Expanded Abstracts (Submitted).
Wong, M., and Ronen, S., and Biondi, B.L., 2012, Joint imaging with streamer and ocean bottom data. SEG Technical Program Expanded Abstracts, 1–5.
Wong, M., and Ronen, S., and Biondi, B.L., 2012, Imaging with multiples using linearized full-wave inversion. SEG Technical Program Expanded Abstracts, 1–5.
Wong, M., 2013, Handling salt reflection in Least-squares RTM SEG Technical Program Expanded Abstracts, 3921–3925.
Wong, M., B. Biondi, S. Ronen, C. Perkins, M. Merritt, V. Goh, and R. Cook, 2014, Robust least squares RTM on the 3D Deimos ocean bottom node dataset SEG Technical Program Expanded Abstracts, 3086–3091. doi: 10.1190/segam2014-0139.1
Wong, M., B. Biondi, and S. Ronen, 2014, Imaging with multiples using least-squares reverse-time migration. The Leading Edge, 33(9), 970976 doi: 10.1190/tle33090970.1
Huy Le and Stewart A. Levin, 2014, Removing shear artifacts in acoustic wave propagation

in orthorhombic media: SEG Technical Program Expanded Abstracts (Submitted).
Huy Le, Robert G. Clapp, and Stewart A. Levin, 2014, Equivalent accuracy at a fraction of the cost: Overcoming spatial dispersion: SEG Technical Program Expanded Abstracts (Submitted).


STANFORD UNIVERSITY
CIRCULATION MAP 1995–96
SCALE (Approx.) 5.7 in. = 1 mile (1 : 11,000)
1/4 mile
indicates areas under construction
INSET
Hoover Pavilion and Arboretum Children's Center (see INSET at upper left)
DOWNTOWN PALO ALTO
Stanford Shopping Center
Town and Country Shopping Center
Medical Center
Stanford Hospital
Lake Lagunita
Golf Driving Range
Golf Course
THE OVAL
MAIN QUAD
STANFORD STADIUM
ATHLETICS
Sports Plaza
Football Practice Field
Maples Pavilion
Eucalyptus Grove
Mausoleum
Frost Amphitheater
Memorial Church
Hoover Tower
Green Library
Cubberley Auditorium
Memorial Auditorium
Art Gallery
Cantor Arts Center
Bing Concert Hall
Law School
Stern Hall
Wilbur Hall
Kimball Hall
Branner Hall
Lagunita Court
Governor's Corner
Tennis Courts
Stables
Golf Course
Observatory
Artist's Studio
Parking and Transportation info: 723-9362
Marguerite Shuttle info: 723-9362
University Telephone Operator: 723-2300
MAP DESIGN BY STEVEN JURNEY FOR STANFORD TRANSPORTATION PROGRAMS
REVISED JULY 1995 (rev. 2.0.1)
Bike route to Menlo Park
Bike route to Palo Alto
Marguerite Shuttle Stop
PA Train Station & Transit Center
Hoover Pavilion (Hospital)
Arboretum Children's Center
Welch Road
Campus Drive West
Campus Drive East
Palm Drive
Sand Hill Road
El Camino Real
Embarcadero Rd
Arboretum Rd
Quarry Rd
Nelson Rd
Galvez St
Lasuen St
Museum Way
Bowl Way
Memorial Way
Serra Mall
Santa Teresa St
Escondido Rd
Lagunita Dr
Mayfield Ave
Junipero Serra Blvd
Stock Farm
Oak Rd
Searsville Rd
Los Arboles Ave
Panama St
Lomita Dr
Governor's Ave
Electioneer
Olmsted Rd
Alvarado Row
Roble Hall
Florence Moore Hall
Lou Henry Hoover House
Escondido Village
Escondido Elementary School
Bing Nursery
Kairos Houses
Mirrielees House
Cowell Student Health Center
Cowell Cluster Houses
Pearce Mitchell Houses
Oliver House
Rogers Marguerite Sena House
Bike route to Palo Alto
Bike route to Alpine Rd & Portola Valley
To Interstate 280 via Alpine Rd or Sand Hill Rd
To Interstate 280 via Page Mill Rd
To Foothill Expressway
Embarcadero Freeway to US 101