2012 Project Review
Consortium Project on Seismic Inverse Methods for Complex Structures
May 14-17, 2012
Breckenridge, Colorado

2011-2012 CWP Faculty, Professors Emeriti and Students

Farhad Bazargani  Dave Hale, CWP Director  Simon Luo  Steven Smith
Norman Bleistein  Detchai Ittharat **  Yong Ma  Roel Snieder
Filippo Brogini  Oscar Jarillo  Andrew Muñoz  Mamoru Takanashi *
Pengfei Cai  Nishant Kamath  Nori Nakata  Ilya Tsvankin
Esteban Díaz  Chinaemerem Kanu  Natalya Patrikeeva  Xiaoxiang Wang
Johannes Douma **  Allison Knaak  Francesco Perrone  Tongning Yang
Ashley Fish  Myoung Jae Kwon  Paul Sava
Clément Fleury  Ken Larner  Bharath Shekar
Jeffrey Godwin *  Luming Liang  Satyan Singh

Long-Term Visitors
Michael Behm, Vienna University of Technology
Jyoti Behura, CWP Post-doctoral fellow
Hongxun Pan, SINOPEC
Kaoru Sawazaki, Tohoku University/Japan Society for the Promotion of Science
Kees Weemstra, Swiss Federal Institute of Technology

CWP Staff/Administration
Shingo Sean Ishida, Communications Specialist
Pamela Kraus, Program Assistant
John Stockwell, Jr., Research Assistant
Diane Witters, Writing Consultant

Center for Wave Phenomena
Colorado School of Mines
1500 Illinois Street, Golden, Colorado 80401
+1.303.384.2178  +1.303.273.3478 (Fax)  cwpcsm@dix.mines.edu

www.cwp.mines.edu
## Contents

<table>
<thead>
<tr>
<th>Acknowledgments</th>
<th>i</th>
</tr>
</thead>
<tbody>
<tr>
<td>CWP policy on proprietary material</td>
<td>iii</td>
</tr>
</tbody>
</table>

### 1) Waveform inversion and tomography

**CWP-705** Application of image-guided full waveform inversion to a 2D ocean-bottom cable data set  
Yong Ma, CWP, Jianxin (Jerry) Yuan and Yunqing Shen, Seismic Technology, ConocoPhillips Company and Bin Gong, In-Depth Geophysical  
1

**CWP-706** Full-waveform inversion of multicomponent data for layered VTI media  
Nishant Kamath and Ilya Tsvankin, CWP  
11

**CWP-707** Full waveform inversion with dynamic image warping  
Yong Ma and Dave Hale, CWP  
21

**CWP-708** Linearized wave-equation migration velocity analysis by image warping  
Francesco Perrone and Paul Sava, CWP, Clara Andreoletti and Nicola Bienati, Eni E&P S.p.A.  
33

**CWP-709** Wavefield tomography based on local image correlations  
Francesco Perrone and Paul Sava, CWP, Clara Andreoletti and Nicola Bienati, Eni E&P S.p.A.  
51

**CWP-710** Image-domain wavefield tomography with extended common-image-point gathers  
Tongning Yang and Paul Sava, CWP  
77

**CWP-711** Illumination compensation for image-domain wavefield tomography  
Tongning Yang, CWP, Jeff Shragge, University of Western Australia and Paul Sava, CWP  
93

### 2) Imaging and velocity analysis

**CWP-712** Understanding the reverse time migration backscattering: noise or signal?  
Esteban Díaz and Paul Sava, CWP  
111

**CWP-713** Optimization of time reversal focusing through deconvolution  
T. J. Ulrich, Los Alamos National Laboratory, Roel Snieder and Johannes Douma, CWP  
125

**CWP-714** Multiparameter TTI tomography of P-wave reflection and VSP data  
Xiaoxiang Wang and Ilya Tsvankin, CWP  
131

**CWP-715** Joint migration velocity analysis of PP- and PS-waves for VTI media  
Pengfei Cai and Ilya Tsvankin, CWP  
149
3) Imaging with multiples

(CWP-716) Increasing illumination and sensitivity of reverse-time migration with internal multiples
   Clement Fleury and Roel Snieder, CWP  157

(CWP-717) Focusing the wavefield inside an unknown 1D medium: Beyond seismic interferometry
   Filippo Broggini and Roel Snieder, CWP, Kees Wapenaar, Delft University of Technology  169

(CWP-718) Creating a virtual source inside a medium from reflection data: a stationary-phase analysis
   Kees Wapenaar, Delft University of Technology, Filippo Broggini and Roel Snieder, CWP  175

(CWP-719) Data-driven wavefield reconstruction and focusing for laterally-varying velocity models, spatially-extended virtual sources, and inaccurate direct arrivals
   Filippo Broggini and Roel Snieder, CWP, Kees Wapenaar, Delft University of Technology  183

4) Image processing and data integration

(CWP-720) Newton-Marchenko-Rose Imaging: Image reconstruction based on inverse scattering
   Jyoti Behura, CWP, Kees Wapenaar, Delft University of Technology and Roel Snieder, CWP  197

(CWP-721) Fault surfaces and fault throws from 3D seismic images
   Dave Hale, CWP  205

(CWP-722) Unfaulting and unfolding 3D seismic images
   Simon Luo and Dave Hale, CWP  221

(CWP-723) Dynamic warping of seismic images
   Dave Hale, CWP  233

(CWP-724) Automatic registration of PP and PS images using dynamic image warping
   Luming Liang and Dave Hale, CWP  245

(CWP-725) Automatically tying well logs to seismic data
   Andrew Muñoz and Dave Hale, CWP  253

5) Attenuation analysis and ultrasonic measurements

(CWP-726) Tensor-guided fitting of subducting slab depths
   Farhad Bazargani and Dave Hale, CWP, Gavin Hayes, U.S. Geological Survey, National Earthquake Information Center  261

(CWP-727) Attenuation analysis for heterogeneous transversely isotropic media
   Bharath Shekar and Ilya Tsvankin, CWP  275
5) Attenuation analysis and ultrasonic measurements (continued)

(CWP-728) *Estimating intrinsic attenuation of a building using deconvolution interferometry and time reversal*
Conrad Newton and Roel Snieder, CWP 283

(CWP-729) *3D synthetic aperture for diffusive fields*
Allison Knaak and Roel Snieder, CWP, Mark Rosenquist, David Ramirez-Mejia and Yuanzhong Fan, Shell International Exploration & Production 293

(CWP-730) *Laser excitation of elastic waves at a fracture*
Thomas Blum and Kasper van Wijk, Boise State University, Roel Snieder, CWP and M. Willis, ConocoPhillips 303

6) Time lapse

(CWP-731) *Analyzing the coda from correlating scattered surface waves*
T. Dylan Mikesell, Kasper van Wijk and Thomas E. Blum, Boise State University, Roel Snieder, CWP and Haruo Sato, Tohoku University 309

(CWP-732) *Estimating near-surface shear wave velocities in Japan by applying seismic interferometry to KiK-net data*
Nori Nakata and Roel Snieder, CWP 315

(CWP-733) *Near-surface weakening in Japan after the 2011 Tohoku-Oki earthquake*
Nori Nakata and Roel Snieder, CWP 331

(CWP-734) *Time-lapse change in anisotropy in Japan’s near-surface caused by the 2011 Tohoku-Oki earthquake*
Nori Nakata and Roel Snieder, CWP 339

(CWP-735) *Estimation of velocity change using repeating earthquakes with different locations and focal mechanisms*
Chinaemerem Kanu and Roel Snieder, CWP 347

(CWP-736) *Feasibility of inverting compaction-induced traveltime shifts for reservoir pressure*
Steven Smith and Ilya Tsvankin, CWP 359

7) Education

(CWP-737) *Teaching graduate students The Art of Science*
Roel Snieder and Ken Larner, CWP and Tom Boyd, Colorado School of Mines 371
Acknowledgments

This project review book is prepared for the sponsors of the Consortium Project at the Center for Wave Phenomena. The Consortium Project provides substantial funding for the overall research and educational program at the Center.

We are extremely grateful for the support of our Consortium sponsors (in alphabetical order):

Aramco Service Co.  
BGP Inc.  
BHP Billiton Petroleum  
BP America Production Co.  
CGGVeritas  
Chevron Energy Technology Co.  
China Petroleum & Chemical Corporation  
Cimarex Energy Co.  
ConocoPhillips  
Devon Energy Corp.  
ENI-AGIP E&P Division  
ExxonMobil Upstream Research  
Geokinetics  
Hess Corp.  
Instituto Mexicano del Petróleo  
ION Geophysical Corp.  
Japan Oil, Gas and Metals National Corp.  
Landmark/Halliburton  
Marathon Oil Corp.  
Nexen Petroleum USA Inc.  
Paradigm Geophysical Corp.  
PDVSA Intevep  
Petrobras America Inc.  
PGS Inc.  
Repsol  
Shell International E & P  
Statoil  
Talisman Energy Inc.  
TGS-Nopec Geophysical Co.  
Total E & P Services  
Transform Software and Services  
WesternGeco, LLC.

The Center also received additional financial support, research grants, fellowships and scholarships from the following agencies and organizations (in alphabetical order):

- BP America  
- ConocoPhillips  
- ENI  
- ExxonMobil  
- NASA  
- The Petroleum Institute, Abu Dhabi  
- Shell  
- U.S. Department of Energy
CWP policy on proprietary material

New printed material produced at the Center for Wave Phenomena with Consortium support is presented to Sponsors before it is released to the general public. We delay general publication by at least 60 days.

During this delay period, Sponsors may use the material within their organizations as they wish. We ask only that all Sponsors respect the rights of other Sponsors and of CWP by not publishing these results externally and independently, in advance of this 60-day delay period. For additional information, please refer to your Consortium Membership Agreement, the paragraph entitled “Sponsor Confidentiality Obligation.”
Application of image-guided full waveform inversion to a 2D ocean-bottom cable data set

Yong Ma\textsuperscript{1}, Jianxin (Jerry) Yuan\textsuperscript{2}, Yunqing Shen\textsuperscript{2} & Bin Gong\textsuperscript{3}

\textsuperscript{1}Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401, USA
\textsuperscript{2}Seismic Technology, ConocoPhillips Company, Houston, TX 77252, USA
\textsuperscript{3}Formerly Seismic Technology, ConocoPhillips Company, Houston, TX 77252, USA; presently In-Depth Geophysical, Houston, TX 77074, USA

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Estimated velocities with the conventional reflection FWI (a) and the image-guided reflection FWI (b).}
\end{figure}

\section*{ABSTRACT}
Image-guided full waveform inversion (IGFWI) was proposed and synthetically tested in Ma et al. (2011). Compared with conventional FWI, IGFWI yields models (e.g., velocities) that make better geologic sense because it takes into consideration the subsurface structures, which can be apparent in migrated seismic images. Moreover, IGFWI converges faster in fewer iterations, especially when reflection energy is used to invert for high-wavenumber details in the model. We test IGFWI on a 2D ocean-bottom cable (OBC) data set. We first use refraction data to update the low-wavenumber component of the model and then proceed to invert for high-wavenumber details with reflection data. In the reflection stage, we incorporate structural constraints in the inversion; as a result, the estimated model makes plausible geologic sense.

\textbf{Key words:} full waveform inversion, image-guided, OBC data

\section{INTRODUCTION}

Full waveform inversion (FWI) (Tarantola, 1984) uses recorded seismic data \( \mathbf{d} \) to estimate parameters of a subsurface model \( \mathbf{m} \) by minimizing the difference between recorded data \( \mathbf{d} \) and synthetic data \( \mathbf{F}(\mathbf{m}) \), where \( \mathbf{F} \) is a forward operator that synthesizes data. In FWI, the objective function often takes an L2 norm: 
\[
E(\mathbf{m}) = \frac{1}{2} || \mathbf{d} - \mathbf{F}(\mathbf{m}) ||^2.
\]
A typical implementation of FWI based on conjugate gradients consists of three steps performed iteratively, beginning with an initial model \( \mathbf{m}_0 \):

(i) compute the gradient of the objective function \( \mathbf{g}_i \) using the adjoint-state method (Tromp et al., 2005);
(ii) search for a step length \( \alpha_i \) in the direction \( \mathbf{h}_i (\mathbf{g}_i, \mathbf{g}_{i-1}) \) (Vigh and Starr, 2008);
(iii) update the model with \( \mathbf{m}_{i+1} = \mathbf{m}_i - \alpha_i \mathbf{h}_i \).

In each iteration, one needs to find a proper
step length decreases the objective function. A quadratic line-search algorithm is performed to find this step length, and the search direction \( h_i \) is defined by the conjugate gradient as

\[
h_0 = g_0 , \quad \beta_i = \frac{g_i^T (g_i - g_{i-1})}{g_{i-1}^T g_{i-1}} , \quad h_i = g_i + \beta_i h_{i-1} . \quad (1)
\]

FWI is a computationally intensive tool. It requires multiple iterations to minimize the data misfit; in each iteration, the cost of both the gradient calculation and the line search is equivalent to the cost of several seismic wavefield simulations and reconstructions, which are especially expensive in 3D.

In addition to its computational cost, FWI also suffers from nonuniqueness because FWI, as a typical inverse problem, is a underdetermined problem. In other words, many different models may yield synthetic data that match recorded data within a reasonable tolerance. This nonuniqueness is caused mainly by local minima in the data misfit function, and the presence of local minima is due to cycle skipping and the nonlinearity in the forward operation \( F(m) \). Strong nonlinearity in reflection FWI makes this local-minima problem more severe (Snieder et al., 1989). Cycle skipping occurs if the time delay between synthetic and recorded data is larger than half a period of the dominating wavelet. In practice, the cycle-skipping problem typically appears because it can be difficult to obtain an adequate initial model that is consistent with unrecoded low frequencies.

Both local-minima and cycle-skipping problems lead to models that poorly approximate the subsurface. To mitigate such problems, multiscale approaches (Bunks, 1995; Sirgue and Pratt, 2004; Boonyasiriwat et al., 2009) have been proposed. The fidelity of multiscale techniques depends fundamentally on the fidelity of low-frequency content in recorded data. In practice, the low frequencies required to bootstrap the multiscale approach may be unavailable.

Ma et al. (2011) propose image-guided FWI (IGFWI) to deal with these problems in a sparse model space, where conventional FWI is posed as a sparse inverse problem. The model \( m \) estimated in conventional FWI is densely sampled for simulating wave propagation. The number of dense samples is far beyond the number of necessary samples needed to geologically explain the model. To reduce the number of required iterations, one may use a sparse representation of the model and reduce the number of model parameters. In inversion, we then should choose as few samples as possible to construct a sparse model \( s \) while still maintaining as many geological features as possible. Between the dense model \( m \) and the sparse model \( s \), we define a linear relationship \( m = Rs \), where \( R \) can take different forms. Various bases (e.g., Fourier bases, wavelet bases, and splines) have been used, typically in tomographic studies, to obtain different sparse models. Among these methods, the representative one is the wavelet transform (Meng and Scales, 1996; Loris et al., 2007; Simons et al., 2011). However, this type of model parameter reduction does not consider the subsurface structures, and as a result, it risks generating geologically non-plausible inversion results. In IGFWI, we implement \( R \) and \( R^T \) with the image-guided interpolation (IGI) (Hale, 2009) and its adjoint operator (Ma et al., 2010), respectively, in order to apply structural constraints derived from migrated images.

In the sparse model space \( s \), we represent FWI as a sparse inverse problem, in which we minimize a new objective function: \( E(s) = \frac{1}{2} ||d - F(Rs)||^2 \). The nonlinear conjugate-gradient method is still valid for us to solve for the sparse model \( s \) iteratively. The line search for this case is performed in a new update direction \( h_s^i \), which is defined as (Ma et al., 2010)

\[
h_s^i = R^T g_i , \quad \beta_i = \frac{(R^T g_i)^T (R^T g_i - R^T g_{i-1})}{(R^T g_{i-1})^T R^T g_{i-1}} , \quad h_s^i = R^T g_i + \beta_i h_s^{i-1} . \quad (2)
\]

In reality, we need a dense model \( m \) to compute synthetic data \( F(m) \) and to fit recorded data \( d \). For this reason, we apply the interpolation operator \( R \) to both sides of equation 2 in IGFWI and thereby interpolate the update direction \( h_s^i \) in equation 2 to obtain an image-guided update direction \( h_m^i \), which corresponds to the dense model space:

\[
h_0^m = RR^T g_0 , \quad \beta_i = \frac{(R^T g_i)^T (R^T g_i - R^T g_{i-1})}{(R^T g_{i-1})^T R^T g_{i-1}} , \quad h_m^i = RR^T g_i + \beta_i h_m^{i-1} . \quad (3)
\]

We then can use the image-guided conjugate-gradient method in equation 3 to update the dense model while maintaining the advantages of sparse inversion.

An implementation of IGFWI based on conjugate gradients consists of four steps performed iteratively, beginning with an initial model \( m_0 \):

(i) compute the gradient \( g_i, R^T g_i, RR^T g_i \);
(ii) search for a step length \( \alpha_i \) in the direction \( h_m^i \);
(iii) update the model with \( m_{i+1} = m_i - \alpha_i h_m^i \);
(iv) remigrate with the updated model and reselect the sparse model based on the remigrated image.

Although the four-step procedure finally gives a dense model solution, it in fact maintains the advantages of sparse inversion. In the sparse inversion, each parameter in the sparse model space \( s \) represents an area in the dense model space \( m \). This sparse representation causes more blocky updates in the model and the blockiness can mitigate the absence of low frequencies in field data. Because the structural features of the subsurface
are taken into consideration as constraints in the sparse inversion, IGFWI generates models that make better geological sense than conventional FWI does. Ma et al. (2011) test IGFWI in the synthetic Marmousi model. In this paper, we test IGFWI on a real 2D OBC data set.

2 DATA

2.1 Field data

Figure 2 shows the source and receiver lines of this 2D data set in a survey map. The entire 2D data set consists of 24 shots and 235 receivers. Shot and receiver spacings are about 500 m and 50 m, respectively. Sources are 5 m below the water surface; receivers are on the sea floor (70 m beneath the water surface). In this OBC survey, both primary and converted waves are collected; however, we only use the P-wave component of the data set in inversion.

Figure 3 shows one P-wave shot gather for the 2D data set; the largest offset is about 12 km. Due to the large offset of the survey, this data set contains significant refraction energy, especially in the far offset. As illustrated in Figure 3, the red line indicates the traveltime of the direct wave; refraction energy arrives earlier than the direct wave. Because of the shallow water depth (70 m), strong surface-related multiples are present in the original data set. However, no processing is done to remove these surface-related multiples. The original data set has a very wide frequency band. A bandpass filter is applied to the original data set to obtain the data between 4 Hz and 20 Hz, shown in Figure 3.

2.2 Source wavelet

Estimation of a source wavelet is an essential step in FWI. In this field data test, the source wavelet for inversion is obtained from a source signature measured in the laboratory. Figure 4a shows the source signature after de-bubbling; its spectrum (Figure 4b) has a much wider range than the effective frequency band (4 – 20 Hz) of the data set. We apply a minimum phase bandpass filter to the source signature and obtain a source wavelet (Figure 5a), with a spectrum (Figure 5b) that matches the frequency band of the data set.
3 METHODOLOGY

3.1 Synthetic data

The survey area is known to have strong vertical transparent isotropy (VTI). Figure 6 shows the initial P-wave velocity model that is produced by traveltime tomography and used in full waveform inversion. Figure 7a and 7b show 1D profiles of the Thomsen parameters $\epsilon$ and $\delta$, respectively. These anisotropic parameters are not exact because they are borrowed from another field that is nearby.

To honor the strong VTI effect in wave propagation, we solve an acoustic VTI wave equation (Chu et al., 2011) that uses the pseudo-acoustic approximation (Alkhalifah, 2000) to simulate the synthetic data. One example of the 2D acoustic VTI wave equation contains two coupled partial differential equations (Du
et al., 2008):

\[
\frac{\partial^2 p}{\partial t^2} = v_{px}^2 \frac{\partial^2 p}{\partial x^2} + v_{pz}^2 \frac{\partial^2 q}{\partial z^2},
\]

(4)

and

\[
\frac{\partial^2 q}{\partial t^2} = v_{pn}^2 \frac{\partial^2 p}{\partial x^2} + v_{pz}^2 \frac{\partial^2 q}{\partial z^2},
\]

(5)

where \(p\) and \(q\) are P- and S-waves, respectively; \(v_{px}\) (Figure 6) is the P-wave velocity along the symmetry axis; \(v_{px} = v_{pz} \sqrt{1 + 2\epsilon}\), and \(v_{pn} = v_{pz} \sqrt{1 + 2\delta}\).

In this study, we only care about the P-wave. Figure 8 shows the synthetic P-wave data simulated with the initial velocity and the Thomsen parameters. If we ignore the VTI anisotropy by using an isotropic acoustic wave equation, the synthetic data has more significant traveltime error in the far offset.

### 3.2 Refraction FWI + reflection FWI

Because significant refraction energy is available in the far offset, we can first use the reflection energy to update the velocity model. In this stage of FWI, we only match seismograms that arrive earlier than the direct wave, indicated by the red line in Figure 3. After updating low-wavenumber components of the velocity with refractions, we then use the reflection information in the data set to update high-wavenumber details of the model. The reflection data that we use for inversion is within the red triangle region, where the maximum zero-offset time is approximately 3 s.

Conventional FWI is performed to implement the inversion with refraction data. When proceeding to the inversion with reflection data, we use IGFWI and compare IGFWI results with conventional FWI results. Both conventional FWI and IGFWI are implemented with a time-domain approach, which is equivalent to inverting the entire frequency band (4 – 20 Hz) simultaneously in the frequency domain.

### 4 INVERSION RESULTS

#### 4.1 Inversion with refraction data

Figure 9a displays the estimated velocity model after 5 iterations of refraction FWI and the difference (b) between this estimation and the initial velocity.

#### 4.2 Inversion with reflection data

In order to allow FWI to update the velocity model with high-wavenumber details, we need to use reflection data. In this stage of reflection FWI, we employ the refraction-FWI-updated velocity model (Figure 9a) as
Figure 10. Gradient $\mathbf{g}$ of the objective function in reflection FWI. The velocity for computing this gradient is the refraction-FWI updated model in Figure 9a.

Figure 11. Estimated velocity (a) after 5 iterations of reflection FWI and the difference (b) between this estimation and the initial velocity.

Figure 12. Structure features (a) indicated by tensor fields (ellipses) and a sparse model represented by dots (b).

4.2.1 Conventional FWI

We first compute the gradient of the objective function using the adjoint-state method. In other words, the gradient is achieved by performing a reverse-time migration (RTM) of the data residual $\mathbf{d} - \mathbf{F} (\mathbf{m}_i)$ with the current velocity model $\mathbf{m}_i$. Figure 10 shows an example of the gradient of the objective function in the reflection FWI. This gradient is computed with the velocity (Figure 9a) that is estimated by conventional FWI with the refraction data. We then use a quadratic line search method to update the velocity model. Figure 11a shows the velocity updated by conventional FWI, which employs the reflection data in 5 iterations. The difference between the reflection-FWI-updated velocity and the initial velocity is shown in Figure 11b. Unlike the refraction inversion (Figure 9b), the inversion with reflection data updates the velocity model with more details by generating high-wavenumber components, as illustrated in Figure 11b. However, the high-wavenumber details are contaminated by many artifacts that are not geologically sensible.

4.2.2 Image-guided FWI

The IGFWI procedure is similar as that of conventional FWI because the nonlinear conjugate-gradient method and the quadratic line search algorithm are the same. The difference is that, instead of computing the gradient $\mathbf{g}_i$, IGFWI computes the projected gradient $\mathbf{R}^T \mathbf{g}_i$ and the image-guided gradient $\mathbf{RR}^T \mathbf{g}_i$. In order to do this, we must first know the structural information of the subsurface and then construct a sparse model according to the structural features.

Figure 12a displays an RTM image of the 2D OBC data, which is migrated with the initial velocity model. On top of the migrated image, ellipses illustrate the structural features of the subsurface, e.g., coherence, orientation. For this field data example, the subsurface structure is horizontally coherent in most areas. Figure 12b shows a sparse model space that is constructed
with the structure-constrained selection scheme (Ma et al., 2011).

Using the structural constraints and the sparse model, we compute the image-guided gradient $RR^Tg$, shown in Figure 13. This image-guided gradient is achieved in two steps. We first apply the adjoint image-guided interpolation operator $R^T$ to the gradient $g$, and obtain the projected gradient $R^Tg$. This step projects (gathers) the gradient information to the chosen sparse model space. We then apply the image-guided interpolation operator $R$ to the projected gradient $R^Tg$, and get the image-guided gradient $RR^Tg$. This step interpolates (scatters) the projected gradient from the sparse to the dense model space. Compared to the regular gradient (Figure 10), the image-guided gradient (Figure 13) contains fewer artifacts and better honors the structural features. Moreover, the gather-scatter process produces low wavenumber components in the image-guided gradient.

The nonlinear conjugate-gradient method (equation 3) takes the projected gradient $R^Tg$ and the image-guided gradient $RR^Tg$, to compute the update direction $h$, employed in the subsequent quadratic line search. Figure 14a shows the velocity updated by the image-guided FWI, which uses the reflection data in 5 iterations. The difference between the reflection-IGFWI updated velocity and the initial velocity is shown in Figure 14b.

### 4.3 Discussion

Because of limited data is used in inversion, velocity update mainly occurs in the area above 3 km. Figure 15 compares the initial velocity, refraction-FWI velocity, conventional reflection-FWI velocity, and the image-guided reflection-FWI velocity. The velocity in Figure 15d, which is obtained after image-guided reflection FWI, shows significantly improvement by removing geologically non-interpretable artifacts from the conventional reflection-FWI model (Figure 15c). Meanwhile, image-guided reflection FWI maintains the high-wavenumber details.

In this field data test, the synthetic data takes into account of VTI anisotropy, but the anisotropic parameters may not be exact because they are estimated for another nearby field. Therefore, the estimated velocity model may have errors to trade off the ambiguity between the velocity itself and the anisotropy parameters.

RTM is used to test the inversion results. Figure 16 displays the RTM images of the 2D OBC data set, which is migrated with the initial velocity, refraction-FWI velocity, conventional reflection-FWI velocity, and image-guided reflection-FWI velocity, respectively. In the circle-highlighted area of Figure 16 have gas clouds been discovered. Despite the existence of gas clouds, rock layers should maintain the structural continuity. However, broken structures are observed in the highlighted area of Figures 16a and 16b. In contrast, Figures 16c and 16d show more interpretable coherent structures, especially in the highlighted gas cloud area. Moreover, the migrated image in Figure 16d, which is done with the image-guided reflection-FWI velocity, contains less artifacts than Figure 16c.

Each iteration of the IGFWI is more expensive than one iteration of conventional FWI as steps (i) and (iv) bring additional cost. Fortunately, the cost of applying the interpolation and its adjoint operator or reselecting a sparse model is negligible compared to the cost of forward modeling or reverse time migration. Also, it is not necessary to reselect the sparse model in every iteration,
Figure 15. Zoom-in views of initial model (a), refraction updated model (b), reflection updated model (c), and image-guided reflection updated model (d).
Figure 16. RTM with initial model (a), refraction-FWI model (b), reflection-FWI model (c), and image-guided reflection-FWI model (d).
and therefore by applying the last step only in selected iterations, we can further reduce the cost.

5 CONCLUSION

We have demonstrated the capability of IGFWI for estimating velocity models with a 2D OBC data set. The refraction and reflection data are sequentially used to update the velocity model. Compared to conventional reflection FWI, IGFWI with reflection data, which essentially solves a sparse inverse problem, generates velocity models that make better geological sense. This improvement is due to the fact that IGFWI uses the subsurface structures, which are extracted from the migrated seismic image, to constrain the inversion in the sparse model space. We compared RTM images of the 2D OBC data using velocity models that are updated by conventional FWI and IGFWI. With the IGFWI-estimated velocity, RTM images show more interpretable coherent structures in the complex area, known to contain low-velocity gas clouds.

6 ACKNOWLEDGMENTS

Authors thank ConocoPhillips for the permission to publish this work, which was finished during Yong Ma’s internship with the Seismic Technology group at ConocoPhillips in summer 2011. Yong wants to thank Yanqing Zeng for his generous help to make RTM images in this report. Yong is grateful to Chunlei Chu and Yi Wang for the reference to the acoustic VTI modeling that is employed in this study. Yong also wants to thank Kirk Wallace for his help to generate common-image gathers, which are used intermittently to QC the inversion.

REFERENCES

Full-waveform inversion of multicomponent data for layered VTI media

Nishant Kamath & Ilya Tsvankin
Center for Wave Phenomena, Colorado School of Mines

ABSTRACT

Although full-waveform inversion (FWI) has shown significant promise in reconstructing heterogeneous velocity fields, most existing methodologies are limited to acoustic models. We extend FWI to multicomponent (PP and PS) data from anisotropic media, with the current implementation limited to a stack of horizontal, homogeneous VTI (transversely isotropic with a vertical symmetry axis) layers. The algorithm is designed to estimate the interval vertical P- and S-wave velocities ($V_P^0$ and $V_S^0$) and Thomsen parameters $\epsilon$ and $\delta$ from long-spread PP and PSV reflections. The forward-modeling operator is based on the anisotropic reflectivity technique, and the inversion is performed in the time domain using the gradient (Gauss-Newton) method. To build the initial model, we perform nonhyperbolic semblance analysis, which yields the zero-offset traveltimes and effective NMO velocities of PP- and PS-waves along with the anellipticity parameter $\eta$. Then the interval parameters are obtained either from Dix-type equations or velocity-independent layer stripping.

1 INTRODUCTION

Transversely isotropic media with a vertical axis of symmetry (VTI) are described by the vertical P- and S-wave velocities, $V_{P0}$ and $V_{S0}$, and the Thomsen parameters $\epsilon$, $\delta$, and $\gamma$. However, traveltime analysis of PP-wave reflection data typically yields just the P-wave normal-moveout velocity $V_{nmo,P}$ and anellipticity coefficient $\eta$ (Alkhalifah and Tsvankin, 1995):

$$V_{nmo,P} = V_{P0} \sqrt{1 + 2\delta}, \quad \eta = \frac{\epsilon - \delta}{1 + 2\delta}. \quad (1)$$

In moveout inversion, $\eta$ is often replaced with the P-wave horizontal velocity:

$$V_{hor,P} = V_{P0} \sqrt{1 + 2\epsilon}. \quad (3)$$

Tsvankin and Thomsen (1995) show that all four parameters responsible for propagation of P- and SV-waves ($V_{P0}$, $V_{S0}$, $\epsilon$, and $\delta$) can be obtained from long-spread PP- and SS-wave traveltimes. Shear waves, however, cannot be excited in offshore surveys and are seldom generated on land. Therefore, here we consider joint inversion of PP-waves and converted PSV modes. The replacement of pure SS reflections with PS-waves, however, complicates velocity analysis because even long-spread traveltimes of PP- and PS-waves are insufficient for constraining the interval parameters $V_{P0}$, $V_{S0}$, $\epsilon$, and $\delta$ of layer-cake VTI media (Grechka and Tsvankin, 2002).

Here, we examine the feasibility of reconstructing layered VTI models in depth using full-waveform inversion of PP and PS data. FWI can be performed either in the time domain (Kolb et al., 1986; Gauthier, 1986; Mora, 1987; Bunks et al., 1995) or frequency domain (Song and Williamson, 1995; Song et al., 1995; Pratt, 1999; Pratt and Shipp, 1999). It is typically based on gradient estimation by zero-lag crosscorrelation of the source and residual receiver wavefields, as described in Tarantola (1984). Most existing algorithms are designed for
isotropic, acoustic media, and operate primarily with diving waves. However, when the medium is anisotropic, it is highly beneficial to combine PP-wave data with PS- or SS-waves, which requires employing elastic models. Taking elasticity into account also makes it possible to properly model reflection amplitudes.

Plessix and Rynja (2010) implement FWI for VTI media in the acoustic approximation to invert for the NMO velocity \( V_{\text{nmo},P} \), anellipticity parameter \( \eta \), and \( \delta \). Plessix and Cao (2011) invert phase information contained in diving waves and near-offset reflections for the long-wavelength components of the NMO (\( V_{\text{nmo},P} \)) and horizontal (\( V_{\text{hor},P} \)) velocities. Single- and multi-parameter FWI for VTI media is performed by Ghoshami et al. (2011). In the former case, they estimate only one velocity parameter (\( V_{\text{hor},P} \), \( V_{\text{nmo},P} \), or \( V_{\text{hor}} \)) while the long-wavelength variations of \( \epsilon \) and \( \delta \) are fixed at the correct values. They also invert for two velocities (\( V_{\text{hor},P} \) and \( V_{\text{hor}} \)) under the assumption that the long-wavelength spatial variation of \( \delta \) are known. They conclude that the single-parameter inversion provides a good estimate of the unknown velocity, while multiparameter inversion can resolve only one of the velocities.

Chang and McMechan (2009) perform a feasibility study of FWI for a horizontal anisotropic layer sandwiched between isotropic media. In addition to transversely isotropic TI layers with a vertical (VTI) and horizontal (HTI) symmetry axis, they also consider a layer of orthorhombic symmetry. They use multicomponent data to invert for the vertical P- and S-wave velocities, anisotropy parameters, and density of the anisotropic layer as well as for the parameters of the underlying isotropic halfspace. Inversion is performed in two steps: first, reflections from the top of the anisotropic layer are used to estimate its parameters; then, the entire data set is inverted for the parameters of both the anisotropic layer and the halfspace. They conclude that wide-azimuth reflections from both the top and bottom of the anisotropic layer are needed for stable parameter estimation.

Here we develop an FWI algorithm for multicomponent data from more realistic, multilayered anisotropic media. PP and PS reflections from all interfaces are inverted simultaneously, which mitigates downward error propagation through the model. First, we describe application of moveout inversion to building the initial model from just PP-wave moveout or from the combination of PP and PS reflection traveltimes. Then we analyze the Hessian matrix for layered VTI models to identify the parameters constrained by input data acquired for a wide range of spreadlength-to-depth (X/Z) ratios. The inversion algorithm is applied separately to just PP data and to the combination of PP and PS reflections to evaluate the feasibility of building VTI depth models from different input data.

## 2 METHODOLOGY

We generate 2D synthetic PP and PS(PSV) data from a point explosive source with the anisotropic reflectivity method (Mallick and Frazer, 1990). In practice, reflection data are sorted into CMP gathers to minimize reflection-point dispersal. However, our FWI algorithm is applied to shot gathers because it operates with synthetic data from horizontally layered media. The parameters of the first layer (or the overburden) are assumed to be known and fixed at the correct values during the inversion.

### 2.1 Building the initial model

To build the initial model, we employ widely used moveout-inversion techniques. Time processing of PP reflection data in VTI media is fully controlled by the parameters \( V_{\text{nmo},P} \) and \( \eta \), which can be estimated from PP-wave traveltimes. The PP-wave long-spread reflection moveout in a horizontal VTI layer is described by the nonhyperbolic equation of Alkhalifah and Tsvankin (1995):

\[
t^2 = t_{P0}^2 + \frac{x^2}{V_{\text{nmo},P}^2} - \frac{2\eta x^4}{V_{\text{nmo},P}^2 \left[ t_0^2 V_{\text{nmo},P}^2 + (1 + 2\eta) x^2 \right]},
\]

where \( x \) is the offset and \( t_{P0} \) is the PP-wave two-way zero-offset time. The velocity \( V_{\text{nmo},P} \) controls the moveout on conventional spreads, while \( \eta \) is responsible for deviation from hyperbolic moveout in long-spread data.

Equation 4 remains valid for layered VTI media, with \( V_{\text{nmo},P} \) and \( \eta \) becoming effective quantities for the stack of layers above the reflector. If the spreadlength-to-depth ratio X/Z is less than 1.5, the magnitude of nonhyperbolic moveout is insufficient for constraining \( \eta \). For X/Z reaching 1.5-2, equation 4 is used to perform 2D semblance scanning and estimate the effective parameters \( V_{\text{nmo},P} \) and \( \eta \). If the offset range is wide enough to record head waves, \( V_{\text{hor},P} \) can be estimated directly from the head-wave moveout (Tsvankin, 2005). Then the interval velocity \( V_{\text{nmo},P} \) is found from the conventional Dix equation and the interval \( \eta \) from the Dix-type equation given in Tsvankin (2005), Appendix 4B.

Because of the trade-off between the moveout parameters, the error in the effective \( \eta \) can be substantial (Alkhalifah, 1997; Grechka and Tsvankin, 1998). In addition, the errors in the effective parameters are amplified by Dix-type differentiation. A more stable technique for estimating the interval moveout parameters is based on velocity-independent layer stripping (VILS) developed by Dewangan and Tsvankin (2006). VILS mitigates error propagation into the interval parameters even in the presence of correlated noise (Wang and Tsvankin, 2009).

The density \( \rho \) and shear-wave vertical velocity \( V_{S0} \) (if only PP data are available) for the initial model are supposed to be found from well logs. The initial value of \( \delta \) is set to zero, which allows us to find the parameters \( V_{\text{hor},P} \) and \( \epsilon \) from \( V_{\text{nmo},P} \) and \( \eta \).

For multicomponent data, it is necessary to identify the PP and PS (PSV) reflections from the same interfaces (i.e., perform event registration). The interval values of \( V_{\text{nmo},P} \) and \( \eta \) can be calculated from P-wave data as described above. To estimate the effective PS-wave NMO velocity \( V_{\text{nmo},P,PS} \), we apply a 2D semblance scan based on equation 4 to long-spread PS data. In this case, \( \eta \) represents just a fitting parameter, but the equation is sufficiently accurate to constrain \( V_{\text{nmo},P,PS} \).
Full-waveform inversion of multicomponent data for layered VTI media

which replaces $V_{nmo,P}$ in equation 4 (Xu and Tsvankin, 2008). Then the effective NMO velocity $V_{nmo,SV}$ of the pure SS reflection can be found from (Seriff and Sriram, 1991):

$$2t_{PSO}V^2_{nmo,PS} = t_{P0}V^2_{nmo,P} + t_{SO}V^2_{nmo,SV},$$  

(5)

where $t_{PSO}$ and $t_{SO}$ are the zero-offset traveltimes of PS- and SS-waves respectively, so

$$t_{SO} = 2t_{PSO} - t_{P0}.$$  

(6)

The SV-wave NMO velocity is given by (Tsvankin, 2005):

$$V_{nmo,SV} = V_{SO}/\sqrt{1 + 2\sigma},$$  

(7)

where

$$\sigma = \left(\frac{V_{SO}}{V_{nmo,SV}}\right)^2(\epsilon - \delta).$$  

(8)

The NMO velocities $V_{nmo,P}$ and $V_{nmo,SV}$ and the vertical-velocity ratio $V_{P0}/V_{SO}$ = $t_{SO}/t_{P0}$ are generally well-constrained by reflection traveltimes. Grechka and Tsvankin (2002) show that in principle it is possible to calculate all four parameters ($V_{P0}$, $V_{SO}$, $\epsilon$, and $\delta$) from $V_{nmo,P}$, $V_{nmo,SV}$, $V_{P0}/V_{SO}$, and $\eta$. The parameter $\delta$ can be found from

$$1 + 2\delta = \left(\frac{t_{P0}}{t_{SO}}\right)^2 \left(\frac{1}{V_{nmo,SV}/V_{nmo,P}} - 2\eta\right).$$  

(9)

Then, $V_{P0}$ and $\epsilon$ are obtained from equations 1 and 2 and $V_{SO}$ from the ratio $t_{SO}/t_{P0}$. However, small errors in the NMO velocities and $\eta$ lead to large errors in the estimated $\delta$, which propagate into the other VTI parameters and make the results too unstable for accurate model-building (Grechka and Tsvankin, 2002). Still, this approach provides us with an initial model to be updated by full-waveform inversion.

2.2 Inversion algorithm

We perform time-domain inversion of either PP data alone or the combination of PP and PS reflections. The least-squares objective function is defined as:

$$F(m) = \frac{1}{2} \|d_{obs} - d_{calc}(m)\|^2,$$  

(10)

where $d_{obs}$ is the observed data and $d_{calc}(m)$ is the data calculated for a certain model $m$. The model updating is carried out via the Gauss-Newton method, with the gradient vector,

$$(\Delta m) = \left[J^T J \right]^{-1} J^T \Delta d,$$  

(11)

where $J$ is the Fréchet derivative matrix obtained by perturbing each model parameter, $J^T J$ is the approximate Hessian, and $\Delta d$ is the difference between the observed data and those computed for a trial model. Forward modeling is carried out with the anisotropic reflectivity algorithm of Mallick and Frazer (1990), based on the formulation introduced by Fryer and Frazer (1984). The main advantage of this method is that it produces the exact reflected wavefield for horizontally layered media including all multiples and mode conversions. In addition, it is possible to separate the wavefield and model either just PP reflections or both PP and mode-converted PS data.

Since the vertical velocities and anisotropy parameters do not have the same units, it is more convenient to invert for the vertical and NMO velocities. If only PP data are used, each layer is described by the parameters $V_{P0}$, $V_{nmo,P}$, $V_{hor,P}$, $V_{P0}/V_{SO}$, and the density $\rho$. In the case of joint inversion of PP and PS data, we estimate the interval values of $V_{P0}$, $V_{SO}$, $V_{nmo,P}$, $V_{nmo,SV}$, and $\rho$. The initial values of $V_{P0}$ and $V_{SO}$ obtained from PP and PS data can be used to calculate the initial $V_{P0}/V_{SO}$ ratio for the inversion of PP-waves. In practice, if only PP reflections are acquired, the initial $V_{P0}/V_{SO}$ has to be known a priori (e.g., from well logs).

2.3 Amplitude signature of P-waves

The inversion algorithm is designed to fit both amplitudes and phase of the modeled data (as part of the full waveforms) to those of the recorded wavefield. The dependence of traveltimes on the parameters of layered VTI media was discussed above. Next, we briefly describe the influence of anisotropy on reflection coefficients and geometrical spreading of PP-waves.

The PP-wave reflection coefficient at a boundary between VTI halfspaces in the weak-contrast, weak-anisotropy ($|\delta| \ll 1$, $|\epsilon| \ll 1$) approximation is given by (Rüger, 1997, 2002):

$$R = \frac{1}{2} \frac{\Delta Z}{Z} + \frac{1}{2} \left[ \frac{\Delta V_{P0}}{V_{P0}} - \left(\frac{2\sqrt{\rho V_{P0}^2}}{V_{P0}}\right)^2 \frac{\Delta G}{G} + \Delta \delta \right] \sin^2 \theta$$

$$+ \frac{1}{2} \left[ \frac{\Delta V_{P0}}{V_{P0}} + \Delta \delta \right] \sin^2 \theta \tan \theta,$$  

(12)

where $\theta$ is the incidence phase angle, $Z = \rho V_{P0}$ is the PP-wave vertical impedance, and $G = \rho V_{SO}^2$ is the S-wave vertical rigidity modulus. The difference between the parameter $B$ below and above the reflector is denoted by $\Delta B = B_2 - B_1$. The average of the two parameters is denoted by $B = (B_1 + B_2)/2$.

The first term in equation 12 is the normal-incidence reflection coefficient, also known as the AVO intercept, which is equal to the fractional difference between the impedances in the two media. The second term is responsible for amplitude variation near the vertical and is called the AVO gradient. It depends on the jumps in $V_{P0}$ and $G$ across the interface and on the contrast in the parameter $\delta$. Hence, PP-wave reflection data at small offsets ought to be sensitive to those parameters; note that $G$ depends on the shear-wave velocity $V_{SO}$. The third term contributes to the amplitude variation at large offsets and is called the “curvature term.”

The far-field amplitude of the PP-wave excited by a point force in a homogeneous, weakly anisotropic TI medium is given by (Tsvankin, 1995, 2005; Xu et al., 2005):

$$A_P(R, \theta) = \frac{F_w}{4\pi \rho V_{P0}^2 R} \frac{1 - 2(\epsilon - \delta) \sin^2 \theta + \delta \sin^2 \theta}{1 + 2\delta},$$  

(13)

where $F_w$ is the projection of the force $\mathbf{F}$ onto the polarization vector, $R$ is the source-receiver distance, and $\theta$ is the phase angle with the symmetry axis. For small angles $\theta$, the amplitude variation with angle is largely controlled by $\eta \approx \epsilon - \delta$. 

which has important implications for our FWI results. Although equation 13 is derived for a homogeneous medium, it also describes the behavior of the anisotropic geometrical-spread factor in any TI layer crossed by the reflected ray (Tsvankin, 2005).

3 INVERSION RESULTS

First, the FWI algorithm is applied to the simple three-layer model in Figure 1. The top layer is isotropic, and its velocities and density are assumed to be known. The bottom halfspace is also known to be isotropic, but its parameters are estimated by FWI. We perform tests for data with the spreadlength-to-depth ratio X/Z ranging from one to three. For X/Z=1, as mentioned before, $\eta$ cannot be constrained by PP reflection traveltimes, so the initial values of $\epsilon$ and $\delta$ are set to zero. For larger spreads (X/Z=1.5, 2, and 3), inversion is performed with the initial $\delta$ either set to zero, or (when PS data are included) computed from equation 9. For PP data, the initial value of $\epsilon$ is then obtained from $\eta$, and the initial $V_{P0}$ from $V_{\text{ano},P}$ (equations 1 and 2).

The testing shows that the interval parameters $V_{P0}$, $V_{S0}$, $\epsilon$, and $\delta$ can be constrained by FWI, but the inversion is extremely sensitive to the starting model when the data include both PP and PS reflections. When PP and PS data are inverted with the initial $\delta = 0$, the algorithm converges to the correct values only for X/Z=1. For longer spreads, accurate inversion requires calculating $\delta$ from equation 9, even though that equation produces errors reaching 0.6. This is likely due to the shape of the objective function, which causes the inversion for the initial $\delta = 0$ to get trapped in local minima.

To evaluate the sensitivity of the objective function to the model parameters, we perform the eigenvector/eigenvalue decomposition of the Hessian matrix for joint inversion of PP and PS data (Plessix and Cao, 2011). Each component of an eigenvector (called the “direction cosine”) indicates the relative sensitivity of the objective function to the model parameters.

Figure 2 shows that the objective function is most sensitive to the layer thickness $D$ (and hence to $V_{P0}$, since the vertical traveltimes are well-constrained), followed by $V_{S0}$, $V_{\text{ano},P}$, and $V_{\text{ano},SV}$. This result is in agreement with Plessix and Cao (2011), who showed that for small-offset P-wave reflected data, the objective function is highly sensitive to the velocity $V_{P0}$. In contrast, all our tests demonstrate that the objective function is weakly sensitive to density. Although the densities change at every iteration, the inversion gets trapped in local minima. Apparently, the objective function becomes much more complicated with the inclusion of density as an unknown parameter. Hence, in all subsequent tests the interval densities are fixed at the correct values.

Next, we generate only PP data for the same model and invert for the parameters $V_{P0}$, $V_{\text{ano},P}$, $V_{\text{hor},P}$, and $V_{P0}/V_{S0}$ using the same range of spreadlengths. For all values of X/Z, the algorithm converges to the correct parameters even when the initial value of $\delta$ is set to zero. Evidently, the objective...
the objective function is sensitive to the combination of $V_P$ and $D$. Indeed, the exact P-wave geometrical-spreading factor is sensitive to $\eta$, which helps estimate all VTI parameters using full-waveform data. Therefore, FWI of PP reflections can reconstruct the depth scale of simple layer-cake models even without including long-offset data.

For PP data, the eigenvectors associated with the two largest eigenvalues of the Hessian (Figure 3(b)) point equally in the direction of two model parameters ($V_{P0}$ and $D$), and so the objective function is sensitive to the combination of $V_{P0}$ and $D$. The third eigenvector, on the other hand, points almost entirely in the direction of $V_{\text{hor},P}$. As mentioned above, the near-offset P-wave amplitude is influenced by $\eta$, which may help resolve $V_{\text{hor},P}$. The objective function for PP-wave inversion is not as sensitive to the $V_{P0}/V_{S0}$ ratio as it is to $V_{P0}$ and $D$. Indeed, the exact P-wave geometrical-spreading factor in the $0^\circ - 40^\circ$ range typically changes by less than 2-3% for the $V_{P0}/V_{S0}$ ratio varying from 1.73 to 2.2 (Tsvankin, 2005). However, the P-wave AVO gradient (and the P-wave reflection coefficient as a whole) includes the jump in the rigidity modulus $G$ (equation 12), which creates a dependence of the FWI objective function on $V_{S0}$.

When larger offsets are included, the velocity $V_{\text{hor},P}$ (or $\eta$) is well-resolved even in the presence of random noise because it governs the magnitude of nonhyperbolic moveout (Figures 4 and 5). Indeed, for X/Z=2 the eigenvector associated with the largest eigenvalue of the Hessian points almost entirely in the direction of $V_{\text{hor},P}$ (Figure 4(b)). The small errors (up to 0.02) in the inverted parameters $\epsilon$ and $\delta$ (Figure 5) are mostly related to the distortion in the vertical velocity $V_{P0}$.

In our inversion we assign equal weights to the horizontal and vertical displacement components. For P-waves recorded on conventional spreads, the largest eigenvalues of the Hessian have a simpler shape with fewer local minima, if only PP data are included.

Interestingly, for X/Z=1 the inversion yields accurate results despite the absence of PS data. This is an unexpected result since such spreadlengths are not sufficient to constrain the horizontal velocity (or the parameter $\eta$) and, therefore, $\epsilon$ using reflection traveltimes. However, equation 13 indicates that the geometrical-spreading factor near the symmetry axis is sensitive to $\eta$, which helps estimate all VTI parameters using full-waveform data. Therefore, FWI of PP reflections can reconstruct the depth scale of simple layer-cake models even without including long-offset data.
sian associated with the horizontal component \( (H_x) \) are much smaller than those for the vertical component \( (H_z) \) (Figure 6). Hence, as expected, the objective function for P-wave inversion on conventional spreads is more sensitive to the vertical displacement. However, for a longer spread \( (X/Z=2) \), the largest eigenvalues of \( H_x \) and \( H_z \) become comparable (Figure 7(a)). In addition, the largest eigenvalue of \( H_x \) is three times or more the other eigenvalues, and the corresponding eigenvector points in the direction of \( V_{\text{hor},P} \) (Figure 7(b)). Therefore, assigning a larger weight to the horizontal component in the objective function for long spreads may result in a faster convergence toward the correct velocity \( V_{\text{hor},P} \).

Next, we test the algorithm on PP and PS data for a more complicated multilayered VTI model (Figure 8). Again, the parameters of the top layer are fixed at the correct values, and the bottom half-space is known to be isotropic. The results indicate that convergence toward the correct parameters is strongly dependent on the initial model. If the initial value of \( \delta \) is set to zero in each layer (which causes a maximum error in \( \delta \) of 0.25) the inversion of PP and PS data gets trapped in local minima. However, if only PP data are inverted, the objective function apparently has a simpler shape (as was the case for the first model), and the algorithm converges toward the correct parameters. In the remaining tests, we focus on PP-wave inversion because accurate parameter estimation for the horizontally layered VTI models considered here can typically be accomplished without using PS-waves.

We contaminate PP data with band-limited \( (10-25 \text{ Hz}) \) random noise, as before. The eigenvector/eigenvalue decomposition of the Hessian matrix indicates that the objective function is most sensitive to the parameters \( V_{P0}, V_{\text{hor},P}, \) and \( D \) of the shallow VTI layer and to the P-wave velocity in the isotropic layer immediately below it. The influence of the parameters of the deeper layers on the objective function is much weaker.
Full-waveform inversion of multicomponent data for layered VTI media

Explosive source
ISO
ISO halfspace
int. 1
int. 2
int. 3
int. 4
1
2
3
Depth (km)
VTI
ISO
VTI
(1)
(2)
(5)
(4)
(3)

Figure 8. Model with two VTI layers sandwiched between isotropic media. The parameters of layer 1 are \( V_P = 2800 \) m/s, \( V_S = 1400 \) m/s, and \( \rho = 1.8 \) gm/cm\(^3\); for layer 2, \( V_{P0} = 3000 \) m/s, \( V_{S0} = 1652 \) m/s, \( \epsilon = 0.1 \), \( \delta = -0.05 \), and \( \rho = 2.1 \) gm/cm\(^3\); for layer 3, \( V_P = 3400 \) m/s, \( V_S = 1800 \) m/s, and \( \rho = 2.4 \) gm/cm\(^3\); for layer 4, \( V_{P0} = 3700 \) m/s, \( V_{S0} = 2000 \) m/s, \( \epsilon = 0.25 \), \( \delta = 0.1 \), and \( \rho = 2.8 \) gm/cm\(^3\); and for the bottom halfspace, \( V_P = 4300 \) m/s, \( V_S = 2200 \) m/s, and \( \rho = 3.1 \) gm/cm\(^3\).

When the spreadlength is equal to the depth of the bottom of the model, the spreadlength-to-depth ratio for the bottom of shallow VTI layer \( (X/Z_4) \) is close to 2.2. Then the parameters of that layer are well-constrained (Figure 9), but there are significant errors for the deeper VTI layer (Figure 10).

However, for longer spreads (the ratio \( X/Z_4 \) for the bottom of the model is equal to two), the parameters of the deeper VTI layer are accurately resolved (Figure 12). Therefore, when data include sufficiently long offsets (which may not be typical in practice), it is possible to invert for \( V_{P0}, V_{S0}, \epsilon, \) and \( \delta \) with only PP-waves. Even in the presence of band-limited random noise with a signal-to-noise ratio of 14, the error in \( V_{P0} \) for the deeper VTI layer is less than 2.2% for \( X/Z_4 = 2 \), and the errors in \( \epsilon \) and \( \delta \) do not exceed 0.03. Per-

Figure 9. Parameters of layer 2 from the model in Figure 8 after each iteration. The input data include PP reflections; the spreadlength-to-depth ratio for the bottom of layer 2 is \( X/Z_2 = 2.2 \) (for the bottom of the model, \( X/Z_4 = 1 \)). The data are contaminated with band-limited (10-25 Hz) random noise; the signal-to-noise ratio is 14.

Figure 10. Parameters of layer 4 from the model in Figure 8 after each iteration. The input data include PP reflections; the spreadlength-to-depth ratio for the bottom of the model, \( X/Z_4 = 1 \). The data are contaminated with band-limited (10-25 Hz) random noise; the signal-to-noise ratio is 14.

Figure 11. Parameters of layer 2 from the model in Figure 8 after each iteration. The input data include PP reflections; the spreadlength-to-depth ratio for the bottom of the model, \( X/Z_2 \approx 4.5 \) (for the bottom of the model, \( X/Z_4 = 2 \)). The data are contaminated with band-limited (10-25 Hz) random noise; the signal-to-noise ratio is 14.

Figure 12. Parameters of layer 4 from the model in Figure 8 after each iteration. The input data include PP reflections; the spreadlength-to-depth ratio for the bottom of the model \( X/Z_4 = 2 \). The data are contaminated with band-limited (10-25 Hz) random noise; the signal-to-noise ratio is 14.
forming isotropic FWI for a VTI media, on the other hand, can lead to depth stretching or overestimation of velocity in the deeper layers (Gholami et al., 2011).

4 CONCLUSIONS

It is well known that the depth scale of horizontally layered VTI models is not constrained by reflection traveltimes of PP- and PS-waves, even if long-spread data are acquired. Here, we show that the interval vertical P- and S-wave velocities and anisotropy parameters $\epsilon$ and $\delta$ of layer-cake VTI media can be estimated by full-waveform inversion of PP and PS reflection data.

The gradient-based inversion algorithm operates in the time domain with PP reflections or with the combination of PP-waves and mode-converted PS-waves. Modeling is carried out with the anisotropic reflectivity method, which generates the exact multicomponent wavefield for 1D anisotropic media. The initial model for FWI is obtained from nonhyperbolic moveout inversion followed by kinematic layer stripping. It should be emphasized that our FWI algorithm estimates the parameters of all layers simultaneously to mitigate downward error propagation.

First, we examine the inversion for a single VTI layer sandwiched between isotropic layers. If the densities are fixed at the correct values, the parameters $V_P^0$, $V_S^0$, $\epsilon$, and $\delta$ are well-constrained by PP-waves alone. Interestingly, PP data help resolve the interval parameters even for conventional spreadlengths ($X/Z=1$). While PP-wave small-offset traveltimes are controlled by $V_P^0$ and $\delta$, the geometrical-spread factor near the symmetry axis depends on $\eta \approx \epsilon - \delta$. Still, it might be beneficial to use multicomponent data (PP and PS) if the level of noise is high.

We also apply FWI to noise-contaminated data from a multilayered VTI model. The sensitivity of the objective function to the interval parameters decreases for the deeper layers. However, if the ratio $X/Z$ for the bottom of the deepest VTI layer reaches two, its parameters can be obtained from the inversion of PP data.

The analysis performed here for stratified VTI media can be generalized for vertical symmetry planes of azimuthally anisotropic models (e.g., orthorhombic). However, geometrical spreading in the symmetry planes of orthorhombic media is influenced by azimuthal velocity variations and has to be modeled with a 3D algorithm. Our inversion algorithm based on the gauss-Newton method cannot be directly extended to laterally heterogeneous media. Still, the results of 1D inversion provide useful insights for designing the inversion operator capable of handling more complicated heterogeneous structures.

5 ACKNOWLEDGMENTS

We are grateful to the members of the A(nisotropy)-Team of the Center for Wave Phenomena (CWP), Colorado School of Mines, for fruitful discussions. We would also like to thank Jyoti Behura (CWP) for participating in our numerous brainstorming sessions.

REFERENCES


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.


Rüger, A., 1997, P-wave reflection coefficients for trans-
Song, Z., and P. R. Williamson, 1995, Frequency-domain acoustic-wave modeling and inversion of crosshole data; part I, 2.5-D modeling method: Geophysics, 60, 784–795.
Full waveform inversion with dynamic image warping

Yong Ma & Dave Hale
Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401, USA

Figure 1. Velocities estimated by conventional reflection FWI (a) and the hybrid FWI with dynamic image warping (b). The correct depth of the deep reflector is at 1.4 km.

ABSTRACT

Full waveform inversion (FWI) can generate high-resolution subsurface models, but often suffers from an objective function with local minima caused mainly by an absence of low frequencies in seismograms. These local minima cause cycle skipping when the initial model for FWI is far from the true model. To avoid cycle skipping, traveltime inversion is often used to compute initial models for FWI.

We propose to incorporate the merits of traveltime inversion in FWI. We use dynamic image warping (DIW) to measure the traveltime misfit between recorded data and synthetic data. When compared with correlation-based techniques often used in traveltime inversion, DIW reduces errors in estimates of time shifts caused by cycle skipping. FWI with DIW then uses these time shifts to mitigate the problem of local minima in the objective function. FWI with DIW is then a hybrid method for inversion that combines the benefits of both conventional FWI and traveltime inversion.

Key words: full waveform inversion, traveltime inversion, dynamic image warping

1 INTRODUCTION

Full waveform inversion (FWI) (Tarantola, 1984; Pratt, 1999) and reflection traveltime tomography (Stork, 1992; Woodward, 1992; Vasco and Majer, 1993; Zelt and Barton, 1998) are two types of methods that geophysicists use to estimate subsurface parameters, such as seismic velocity models. Both have their own advantages and disadvantages.

FWI uses recorded seismic data \( \mathbf{d}_0(x, t) \) to estimate parameters of a subsurface model \( \mathbf{m} \), by minimizing the difference between recorded data \( \mathbf{d}_0 \) and synthetic data \( \mathbf{d}(x, t) \equiv \mathcal{F}(\mathbf{m}) \), where \( \mathcal{F} \) is a forward operator that synthesizes data. FWI is usually formulated as an optimization problem (Tarantola, 2005), in which we consider the minimization of a multivariate nonlinear objective function \( E(\mathbf{m}) \). The objective function can have various forms (Crase et al., 1990), e.g., L2 norm (Tarantola, 2005).
norm (Guitton and Symes, 2003). In this paper, we use L1 norm (Loris et al., 2007), and Huber norm (Guitton and Symes, 2003). In this paper, we use a least-squares form:

\[ E(m) = \frac{1}{2} \| \mathbf{d}(x,t) - \mathbf{d}_0(x,t) \|^2, \]  

where \( \| . \| \) denotes an L2 norm.

In principle, all information in recorded seismic waveforms should be taken into account in minimizing this difference in FWI. FWI comprehensively minimizes differences in traveltimes, amplitudes, converted waves, multiples, etc. between recorded and synthetic data. This all-or-nothing approach distinguishes FWI from other methods, such as traveltime tomography that focuses on only traveltime differences. Because FWI uses all of the information in the data, it can update the model with high-resolution details. However, FWI often suffers from problems such as local minima and cycle skipping (Snieder et al., 1989). These problems often arise when the initial model is poor or the recorded data do not have sufficient low frequencies.

Ray-based or wave-equation traveltime tomography uses only traveltime differences \( \Delta \tau \) between recorded data \( \mathbf{d}_0 \) and synthetic data \( \mathbf{d} \). Traveltime tomography can also be solved as a least squares optimization problem, where we minimize the following objective function:

\[ E(m) = \frac{1}{2} \| \Delta \tau \|^2. \]  

As noted in Luo and Schuster (1991), traveltime tomography is constrained by a high-frequency assumption about the data. This assumption causes traveltime tomography to fail to estimate velocity variations with wavelengths similar to or smaller than those in the source wavelet. As a result, compared to FWI, traveltime tomography only updates a velocity model with large-scale variations. Nevertheless, large-scale model update is important for FWI. When recorded data lack low frequencies, FWI cannot estimate large-scale components of a velocity model and can possibly lead inversion to local minima. Therefore, it is a common practice to use traveltime inversion to compute initial models for FWI.

Common methods for measuring the traveltime difference \( \Delta \tau \) are based on correlations of recorded data \( \mathbf{d}_0 \) and synthetic data \( \mathbf{d} \) (Luo and Schuster, 1991; van Leeuwen and Mulder, 2010). Unfortunately, correlation-based methods also suffer from the cycle-skipping problem. In this paper, we use dynamic image warping (DIW) (Hale, 2012) to measure the traveltime misfit. DIW, as a global optimization, can avoid cycle skipping. We then design FWI with DIW by using the DIW-estimated traveltime misfit in FWI. With the traveltime misfit, we propose two methods: traveltime dominant FWI and hybrid FWI. Synthetic tests demonstrate that traveltime dominant FWI avoids cycle skipping and hybrid FWI takes the benefits of both traveltime dominant FWI and conventional FWI.

![Figure 2. Top: a synthetic seismic trace (solid) and a shifted trace (dotted). Bottom: true traveltime misfit (dotted), misfit found by crosscorrelation (dashed) and dynamic warping (solid).](image-url)
Local crosscorrelation, in this example, fails to detect the shifts; the estimated shifts (dashed line) are significantly deviated from the true shifts (dotted line) due to cycle skipping.

Another shortcoming of this correlation-based approach is that it works only if the source spectra of the recorded and the synthetic data are the same (de Hoop and van der Hilst, 2005). Therefore, van Leeuwen and Mulder (2010) modified the correlation-based approach with a weighted norm to address this shortcoming.

2.2 Dynamic image warping

To deal with the cycle-skipping problem, we choose dynamic image warping (DIW) (Hale, 2012) as a method to estimate the traveltime misfit. DIW obtains the traveltime misfit \( u(x,t) \) between the recorded data \( d(x,t) \) and the synthetic data \( d_0(x,t) \) by finding a shift field \( s(x,t) \) that solves the following constrained optimization problem:

\[
  u(x,t) = \arg \min_{s(x,t)} D(s(x,t)) ,
\]

where

\[
  D(s(x,t)) = \|d(x,t) - d_0(x,t + s(x,t))\|^2 ,
\]

subject to the constraint

\[
  \left| \frac{\partial u(x,t)}{\partial t} \right| \leq 1 .
\]

The constraint in equation 7 is adapted from the simplest slope constraint of Sakoe and Chiba (1978). This constraint ensures that the traveltime misfit \( u(x,t) \) neither decreases nor increases too rapidly in the \( t \) axis.

Unlike the local crosscorrelation-based approach, DIW is a global optimization approach that can avoid cycle skipping. Despite the added noise, DIW correctly detects the traveltime misfit between \( f[i] \) and \( g[i] \), as indicated by the solid line in the lower panel of Figure 2.

3 FWI WITH DYNAMIC IMAGE WARPING

After we obtain the traveltime misfit \( u(x,t) \) between the recorded data \( d(x,t) \) and the synthetic data \( d_0(x,t) \), we can use this misfit in FWI. To do this, we need to formulate objective functions that are minimized (or maximized) when the misfit \( u(x,t) \) is zero, indicating the reflections in the recorded data \( d(x,t) \) and in the synthetic data \( d_0(x,t) \) are well aligned.

3.1 FWI implementation

A typical implementation of FWI based on conjugate gradients consists of three steps performed iteratively, beginning with an initial model \( m_0 \). In the \( i^{th} \) iteration, we

(i) compute the gradient \( g_i \equiv g(m_i) = \frac{\partial E(m_i)}{\partial m_i} \) with the adjoint-state method (Tromp et al., 2005);
(ii) search for a step length \( \alpha_i \) in the conjugate gradient direction \( h_i(g_i,g_{i-1}) \) (Vigh and Starr, 2008);
(iii) update the model with \( m_{i+1} = m_i - \alpha_i h_i \).

The adjoint-state method computes the gradient as

\[
  g_i = F^*(A(m_i)) ,
\]

where \( F^* \) is the adjoint of the forward operator \( F \), and \( A(m_i) \) is the adjoint source. The adjoint source for FWI is defined as

\[
  A(m) = \frac{\partial E(m)}{\partial d} ,
\]

which can be rewritten as

\[
  A(m) = d - d_0 .
\]

Therefore, the gradient \( g_i \) in FWI becomes \( F^*(d - d_0) \), indicating a reverse time migration of the data residual \( d - d_0 \).

3.2 Traveltime dominant FWI

The most straightforward way of using the misfit \( u(x,t) \) is analogous to conventional traveltime tomography, in which, we aim to solve a least squares inverse problem:

\[
  E(m) = \frac{1}{2}\|u(x,t)\|^2 .
\]

If we use the three-step iterative approach outlined above to solve this optimization problem, we first must compute the gradient of the objective function in equation 11 with respect to the model parameters \( m \). The adjoint-state method is an efficient way to compute the gradient, and for this objective function, the adjoint source is

\[
  A(m) = u \frac{\partial u}{\partial d} .
\]

However, this adjoint source is difficult to compute because we have no close-form expression for \( \partial u/\partial d \).

To avoid the difficulty in evaluating \( \partial u/\partial d \), we choose another objective function that involves both the traveltime misfit \( u(x,t) \) and the synthetic data \( d(x,t) \):

\[
  E(m) = \frac{1}{2}\|u(x,t)d(x,t)\|^2 .
\]

When the misfit \( u(x,t) \) is zero, the objective function in equation 13 still attains a minimum. Because \( \|d(x,t)\|^2 \) cannot be zero, equation 13 is dominated by the traveltime misfit \( u(x,t) \). We then refer to the inverse problem implied by equation 13 as traveltime dominant FWI.

The adjoint source in this case becomes

\[
  A(m) = ud^2 \frac{\partial u}{\partial d} + u^2 d .
\]
which we approximate as
\[ A(m) \approx u^2 d. \] (15)

The assumption behind this approximation is that the traveltime misfit \( u \), as an implicit function of the synthetic data \( d \), varies slowly so that \( \partial u / \partial d \) is small and we can discard the first term on the right hand side of equation 14. The validity of this assumption has not yet been strictly tested, and it may be wrong.

3.3 Hybrid FWI

The objective function in Equation 13 becomes zero when traveltimes in recorded and synthetic data are well aligned, even though amplitudes in synthetic data are incorrect. The traveltime dominant FWI implied by equation 13 primarily uses traveltime information, and as a consequence, solving traveltime dominant FWI implied by equation 13 is similar to performing a traveltime tomography. As discussed in the introduction, traveltime tomography recovers only the large-scale components of a velocity model (Luo and Schuster, 1991). On the other hand, FWI is not limited by traveltime and its objective function uses all the information in the data. Therefore, FWI can recover a velocity model with fine details.

To gain the advantages of both traveltime tomography and FWI, we advocate the following hybrid objective function, which combines objective function for traveltime dominant FWI in equation 13 and conventional FWI in equation 1:
\[ E(m) = \frac{1-c}{2} \| u(x,t) d(x,t) \| ^2 + \frac{c}{2} \| d(x,t) - d_0(x,t) \| ^2, \] (16)

where \( c = e^{-u^2/\sigma^2} \) and \( \sigma \) is a constant that controls the width of this Gaussian. The adjoint source then becomes
\[ A(m) \approx (1-c) u^2 d + c(d - d_0), \] (17)

where we employ the same assumption as in equation 15.

When the traveltime misfit \( u \) is large, \( c \to 0 \) and hybrid FWI in equation 16 becomes traveltime dominate FWI in equation 13; when \( u \) is small, \( c \to 1 \) and hybrid FWI becomes conventional FWI in equation 1. Therefore, the inverse problem posed in equation 16 changes smoothly from traveltime dominant inversion to conventional FWI.

4 EXAMPLES

We test FWI with dynamic image warping for two different synthetic models: one with velocity anomalies and another with a deep reflector.

4.1 Example 1: velocity anomaly

Figures 3a and 3b display true velocity models that contain a high- or a low-velocity anomaly, respectively. The three-layer model without the anomalies shown in Figure 3c is the initial model for testing FWI. In this example, we use 24 shots uniformly distributed on the surface, and 8 Hz Ricker wavelet as the source for simulating wavefields.

Figure 4a displays a shot gather simulated for the model with a high-velocity anomaly, and Figure 4c shows the corresponding shot gather simulated in the

Figure 3. True velocity models that contain a high- (a) and a low-velocity (b) Gaussian anomalies. (c) Shows an initial model for inversion.
Figure 4. Recorded data $d_0$ simulated in the true models with a high-velocity anomaly (a) and a low-velocity anomaly (b). Synthetic data $d$ (c) is computed for the initial velocity model with no anomaly.

The data residuals in Figure 5 indicate that cycle skipping occurs for both the high- and low-velocity anomalies. More quantitatively compelling evidence of cycle skipping is the travelttime misfit $u(x, t)$ between the recorded data $d_0$ and the synthetic data $d$. Dynamic image warping is used to measure these time shifts. Figure 6a displays the absolute time shift caused by the high-velocity anomaly; Figure 6b shows the time shift normalized by the dominant period in the data. In Figure 6b we can observe shifts that are greater than one cycle, indicating the presence of cycle skipping. Like-
Figure 5. Data residuals $d - d_0$ corresponding to the models with high-velocity anomaly (a) and low-velocity anomaly (b).

Figure 6. Traveltime misfit due to the high-velocity anomaly: absolute value (a) and shift (b) normalized by dominant period.

wise, Figure 7 shows the time shifts caused by the low-velocity anomaly and illustrates that cycle skipping occurs in the low-velocity anomaly as well.

Because of cycle skipping in this example, conventional FWI fails to recover neither of these velocity anomalies. Figures 8a and 9a show the conventional FWI velocity estimations for the high- and low-velocity anomalies, respectively. Comparing these estimates to the true velocities in Figures 3a and 3b, we observed that conventional FWI is trapped in local minima and updates the velocity in the direction that is opposite to the correct one.
Figure 7. Traveltime misfit due to the low-velocity anomaly: absolute value (a) and shift (b) normalized by dominant period.

Figure 8. High-velocity anomaly with conventional FWI (a) and traveltime dominant FWI (b).

Figure 9. Low-velocity anomaly estimated with conventional FWI (a) and traveltime dominant FWI (b).
Unlike conventional FWI, traveltime dominant FWI (equation 13) does not compute directly the data difference $d - d_0$; instead, it aims to minimize the traveltime misfit $u(x, t)$ between $d_0$ and $d$. The optimization problem implied by equation 13 avoids the cycle skipping by emphasizing the traveltime misfit $u(x, t)$ in the objective function.

For the high- and low-velocity anomalies, Figures 8b and 9b show respectively the updated models estimated by traveltime dominant FWI. Comparing these estimates with the true velocities shown in Figure 3a and 3b, we see that traveltime dominant FWI updates the velocities in the correct direction and recovers the anomalies reasonably well.

4.2 Example 2: deep reflector

Figure 10a displays a layered velocity model that is the true model in this second example. Figure 10b shows an initial velocity model, and the difference between the true and the initial models is in the depth of the deep reflector, which is misplaced by approximately 125 m. In this example we hope to recover the correct depth of the deep reflector from surface-recorded seismic data. We use 24 shots uniformly distributed on the surface, and a 15 Hz Ricker wavelet is used as the source.

Figures 11a and 11b display shot gathers simulated for the true and initial velocity models, respectively. The difference between the recorded data $d_0$ (Figure 11a) and the synthetic data $d$ (Figure 11b) is the initial data residual $d - d_0$, shown in Figure 12a. Two distinct events appear in the data residual, suggesting cycle skipping caused by the misplacement of the deep reflector. This cycle skipping can also be observed in the traveltime misfit $u(x, t)$ between the recorded data and the synthetic data. Figure 13a shows the traveltime misfit measured by dynamic image warping; Figure 13b plots the normalized shift, indicating the presence of cycle skipping as the maximum normalized shift is greater than one cycle.

Due to cycle skipping, conventional FWI fails to correct the misplaced deep interface, as illustrated in Figure 14a. We computed the synthetic data $d$ with the updated velocity model (Figure 14a) and show the data residual $d - d_0$ in Figure 12b. Compared to the initial residual (Figure 12a), the new residual (Figure 12b) shows smaller magnitude; however, the two events that are caused by cycle skipping become separated farther, indicating that conventional FWI updates the model in a wrong direction.

Figure 14b shows the velocity model updated by traveltime dominant FWI (equation 13) in 10 iterations. This updated model correctly positions the deep reflector while showing a smooth interface. In the updated data residual shown in Figure 12c, we can find only one event, suggesting that the inversion avoids cycle skipping. However, this new residual still contains high amplitudes due to the smooth update. Although equation 13 involves both the traveltime misfit $u$ and the synthetic data $d$, it is the misfit that most significantly controls the objective function. In other words, traveltime dominant FWI uses primarily the traveltime information and therefore performs like a traveltime tomography. This explains why traveltime dominant FWI cannot recover the sharp interface.

We can improve the inversion implied by equation 13 by applying a hybrid FWI formulated in equation 16. Figure 14c displays the velocity model estimated by the hybrid FWI in 10 iterations. Compared with Figures 14a and 14b, Figure 14c not only places the deep reflector at the correct depth, but also recovers the sharp boundary. In early iterations, the hybrid FWI behaves predominantly like traveltime dominant FWI due to a large traveltime misfit; in this stage, the hybrid FWI mainly moves the deep reflector towards the correct depth. As the traveltime misfit decreases, the hybrid FWI gradually becomes conventional FWI;
in this later stage, the hybrid FWI sharpens the reflecting interface. Figure 12d shows the data residual $d - d_0$ corresponding to the hybrid FWI model. The data residual in Figure 12d shows a substantially lower magnitude that is apparent in the other three data residuals.

5 CONCLUSION

We have proposed using DIW as a tool to estimate traveltime misfit between recorded seismic data and synthetic data in FWI. As a global optimization, DIW overcomes the cycle-skipping problem in common traveltime estimations based on local crosscorrelations. We use the DIW-derived traveltime misfit in FWI by defining two objective functions that involve the misfit. Traveltime dominant FWI implied by equation 13 can avoid cycle skipping when the initial model is far from the true one; however, it does not provide high-wavenumber details because it predominantly uses traveltime information in inversion. The hybrid FWI implied by equation 16 combines the benefits of both traveltime dominant FWI and conventional FWI.

The adjoint-state method is used to compute the gradient $\frac{\partial E(m)}{\partial m}$. In order to conveniently compute adjoint sources, we made an assumption that $\frac{\partial u}{\partial d}$ is negligible. This assumption requires further investigation. However, the alternative but useful objective functions may exist with assumptions more valid than those proposed in this paper.

6 ACKNOWLEDGMENTS

This work is sponsored by a research agreement between ConocoPhillips and Colorado School of Mines (SST-20090254-SRA). Yong Ma thanks CWP writing consultant Diane Witters for her help in learning how to improve this manuscript.

REFERENCES


Pratt, R. G., 1999, Seismic waveform inversion in the
Figure 12. Data residuals $d - d_0$ computed for the initial velocity model (a), and for velocity models updated by the conventional FWI (b), traveltime dominant FWI (c), and the hybrid FWI (d).

frequency domain, part 1: Theory and verification in a physical scale model: Geophysics, 64, 888.
——–, 2005, Inverse problem theory and methods for
**Figure 13.** Traveltime misfit detected by DIW: absolute shift (a) and shift (b) normalized by dominant period.

**Figure 14.** Velocities estimated by conventional FWI (a), traveltime dominant FWI (b), and the hybrid FWI (c).


Linearized wave-equation migration velocity analysis by image warping

Francesco Perrone¹, Paul Sava¹, Clara Andreoletti², and Nicola Bienati²

¹ Center for Wave Phenomena, Colorado School of Mines
² eni E&P

ABSTRACT

Seismic imaging by wavefield extrapolation aims to produce images of contrasts in physical parameters in the subsurface, e.g., velocity or impedance. In order to build such images, a background model describing the wave kinematics in the Earth is necessary. In practice, both the structural image and the background velocity model are unknown and have to be estimated from the acquired data. Migration velocity analysis deals with estimation of the background velocity model in the framework of seismic migration and relies on two main elements: the data redundancy and the invariance of the structures with respect to different seismic experiments. Since all the experiments probe the same model, the reflectors must be invariant in suitable domains (e.g., shots or reflection angle); the semblance principle is the conventional tool used to evaluate the invariance of a set of images. Here, we present a velocity error measure that is based on the semblance principle in the shot-domain. This measure is purely kinematic and considers the angle between the structural dip (estimated from the migrated image), and the displacement vector field, which is estimated from images obtained from shots. The maximum constructive interference, i.e., the maximum semblance, requires that the two images superimpose along the reflector slope. If the two images are not in phase along the reflector, then the velocity model is incorrect, and we can measure this discrepancy by estimating the angle between the dip of the reflector and displacement vector field. We derive an expression for the image perturbation that drives a migration velocity analysis procedure based on a linearization of the wave-equation with respect to the model parameters; we then compare our method with the direct implementation based on the difference of images from nearby experiments. The two approaches lead to similar optimization problems. For nearby shots, the image difference has the characteristics of an image perturbation and is robust against cycle skipping. It, however, is sensitive to amplitude patterns and even for the correct velocity model it can produce nonzero velocity updates. Synthetic 2D examples show promising results in retrieving macro velocity models. This methodology can be directly applied to 3D.

Key words: tomography, migration velocity analysis, imaging

1 INTRODUCTION

Seismic imaging aims to construct an image of geologic structures in the subsurface from reflection data recorded at the surface of the Earth. For constructing such an image, one needs the Green functions that describe the wave propagation in the subsurface and hence a model for computing the Green functions. If we assume a linear, acoustic, and constant-density model, the velocity is the only parameter that governs the wave propagation. The correct velocity model is unknown and must be estimated from the data to obtain an accurate image of the reflectors in the subsurface, especially for highly heterogeneous geologic configurations. Complex wave propagation phenomena (for example, multipathing) cannot be correctly handled and the structures in the subsurface cannot be properly imaged, unless the background velocity model is accurately estimated.

Estimating the velocity model from the recorded data is referred to as velocity analysis. The problem is intrinsically nonlinear (because the wave-equation is nonlinear in the velocity parameter) and it is usually formulated as an optimization problem in which the correct velocity model minimizes an objective function. Two different classes of methods for estimating the wave propagation velocity have been dis-
cussed in the literature: waveform inversion (WI) (Tarantola, 1984; Woodward, 1992; Pratt, 1999; Plessix, 2009) and migration velocity analysis (MVA) (Yilmaz, 2001; Sava and Biondi, 2004; Shen and Symes, 2008; Symes, 2008). WI operates in the data domain and iteratively updates the model parameters until the energy of the residual between simulated and recorded data is minimized. The WI objective function is characterized by numerous local minima and a good initial guess of the velocity model is necessary in order to converge to the correct solution. Migration velocity analysis relies on the assumption that the reflectors in the subsurface must be imaged at the same locations by different experiments. When the correct velocity model is used, the similarity between the images constructed for different experiments must be maximum. The similarity of the migrated images (i.e., the semblance principle (Al-Yahya, 1989; Symes, 2008)) constitutes the criterion for designing an objective function that measures the quality of the velocity model.

Evaluating the semblance of a set of images requires constructing common-image gathers, i.e., new images indexed in spatial position and either an extension parameter (e.g., incidence angle, space/time lags (Rickett and Sava, 2002; Sava and Fomel, 2003) or experiments (Soubaras and Gratacos, 2007; Xie and Yang, 2008)). Image gathers allow one to assess the consistency between the velocity model and the recorded data. The construction of image gathers requires the migration of the entire survey or at least a subset of the recorded shots that illuminate the portion of the subsurface we are interested in. Consequently, in order to handle the disk storage requirements and computational cost associated with the analysis, it is necessary to subsample the locations where the gathers are constructed.

In this paper, we present a measure based on the differential semblance criterion (Symes and Carazzone, 1991) for evaluating the correctness of the velocity model in the framework of migration velocity analysis. Our method operates in the image-domain and computes the displacement vector field between two images. The displacement vector field is defined by a warping transformation of one image into the other and is pointwise measured by local crosscorrelations of the two images. The displacement vector field and the estimated structural dip give us tools for extracting a purely kinematic measure of semblance in the image domain without constructing common-image gathers (CIGs).

The motivation for the work is multifold. First, by operating on a shot-by-shot basis, we eliminate the cost associated with construction of image gathers, which requires the migration of the entire dataset or, alternatively, of the shots that illuminate the target area. Consequently, the disk storage is reduced, thus decreasing the complexity and speeding-up the actual migration velocity analysis algorithm. Second, typical CIGs constructed at fixed lateral positions capture the vertical movement of the image point as a function of the extension parameter, although an error in the velocity model can produce a movement in all directions in space. Our technique does not assume a simple vertical shift of the image point and restates the semblance principle in terms of the consistency between the structural information in the image and the apparent displacement of image points as a function of experiments. Finally, the shot-by-shot approach allows one to bootstrap the MVA procedure without recording the whole dataset; thus we can update the velocity model in an iterative fashion with respect to shots.

In the following sections, we describe the semblance principle, we review the wave-equation migration velocity analysis (WEMVA) procedure, and then we introduce our approach for measuring the consistency of the velocity model based on the apparent shift between two images from adjacent shots.

2 THE SEMBLANCE PRINCIPLE

In seismic migration, the velocity model is assumed to be known but in reality it represents the main unknown and has to be estimated from the data. Migration velocity analysis measures the similarity of the different migrated images (which are obtained by exploiting the redundancy of the data) using the semblance principle (Al-Yahya, 1989). The semblance principle relies on the invariance of the subsurface with respect to the seismic experiments: since the model that generates the data is unique and time-invariant, different experiments must image the same structures. The quality of the migration result is assessed by constructing an image cube that collects the images as a function of the spatial coordinates x and experiment R(x, e) or extension parameter R(x, α). The variable e indexes the experiments and can represent the shot number or the ray-parameter in plane-wave migration; the extension parameter α can represent the reflection angle or the subsurface offset. The current practice consists in analyzing fixed spatial locations and considers all the experiments (Figure 1(a)) or all the values of the extension parameter (Figure 1(b)). If the velocity model is correct, the images show invariance along these dimensions in the image cube. This property is true in a kinematic sense: the reflection coefficients (Aki and Richards, 2002), which depend nonlinearly on the incidence angle, are neglected by this methodology. Several choices of domain are available for analyzing the invariance in the image cube, for example the extended image domain (Rickett and Sava, 2002; Symes, 2008; Vasconcelos, 2008), the reflection angle domain (Sava and Fomel, 2003; Biondi and Symes, 2004b), and the experiment domain (Chavent and Jacewitz, 1995; Soubaras and Gratacos, 2007; Xie and Yang, 2008).

The semblance principle can be implemented in different fashions:

(i) In conventional semblance (Taner and Koehler, 1969), the energy of the stack of the migrated images measures the quality of the velocity model; the correct velocity model maximizes that energy;

(ii) In differential semblance (Symes and Carazzone, 1991), the energy of the first derivative along the extension axis measures the correctness of the velocity model; the correct model minimizes the energy.
Both conventional and differential semblance are customarily implemented using gathers, e.g., common-offset or common-reflection angle, in order to analyze the consistency between different results. Inconsistency in the migrated images appears as moveout in the gathers; the moveout is used for estimating the velocity error and for computing the velocity update for the current model. If we consider common-image gathers, we have the situation depicted in Figures 1(a) and 1(b). Since differential semblance considers similarities between adjacent experiments, in principle we can work without gathers in an iterative fashion by considering pairs of experiments (Figure 1(c)). This approach allows one to analyze all the points in the aperture of the two experiments at the same time, instead of constraining the appraisal of the velocity model at specific spatial locations where the CIGs have been constructed.

In this work, we explore an alternative method for measuring semblance in the experiment domain. The experiment index may represent the shot number or the plane-wave ray-parameter, and we consider all points in the aperture shared by adjacent shot-gathers. The semblance principle is implemented in a differential sense by measuring the constructive interference of two images. This point of view has several advantages. The computational cost is reduced since we can analyze all the points in the aperture shared by two experiments at once instead of looking at particular spatial locations where the gathers are constructed. Thus, the resulting velocity update influences a wider portion of the velocity model and potentially speeds up the convergence towards a solution. Differential semblance in the experiment domain is flexible and can be implemented independent of the migration algorithm used for imaging the data. The potential disadvantages are related to the amplitude differences between the two images: two neighboring experiments are likely to supply the same structural image but with different amplitude footprint. If the velocity field that generated the data produces observable changes in amplitude between adjacent experiments (for example, because of shallow lenses), the simple difference between the corresponding images (if small but not zero) is an inaccurate measure of similarity and can lead to spurious velocity updates even if the migration velocity model is correct. We develop a measure that is not so sensitive to the amplitude patterns.

3 IMAGE WARPING AND VELOCITY MODEL INCONSISTENCY

The seismic image forms at the stationary points of the travelt ime between the source and receiver position. When we stack the images from different shots, we evaluate the station ary point with respect to the shot position; specular reflectors build up along the structural slope, and the residual diffractions caused by limited aperture stack out. If the velocity model is incorrect, the stationary point at depth changes with experiment: the different images are not “in phase”, and the interference is not totally constructive, thus reducing the overall amplitude of the stack. If we consider two adjacent experiments (e.g., shots), we can assess the consistency between the corresponding partial images (i.e., images obtained from single shots or plane waves) by evaluating the degree of constructive interference of the migrated events. In other words, we measure the relative displacement of two images. A conventional differential semblance approach requires the migration of the entire survey and compares pairs of adjacent images after a mapping in a suitable domain (e.g., in the angle-domain) at fixed spatial locations; our differential semblance approach compares pairs of images from adjacent experiments at every image point and measures inconsistencies as phase shifts in the image space. Because this methodology includes all points in the aperture at once, it does not require the construction of gathers. The phase shift between the two images is measured as the angle between the dip field estimated from the reflectors in the images and a second vector field that warps one image into the other.

Figure 1. Migrated images can in principle be ordered in a hypercube and indexed by spatial location \( \mathbf{x} \), experiment number \( e \), and extension parameter \( \alpha \). By stacking along one of the axes we reduce the dimensionality of the data and are able to analyze them. For example, we can stack along the extension parameter axis (e.g., reflection angle) and evaluate model accuracy by measuring the semblance at discrete positions in space between different results. Inconsistency in the migrated images appears as moveout in the gatherings, which we usually analyze at fixed spatial locations (b). A third option is to stack over the extension parameter and to analyze all points in the image for two (or a small number of) images obtained from nearby experiments (c). This last solution is what we propose in this paper.
Figure 2. Horizontal reflector for correct velocity model (a); the images obtained from two adjacent shots (red and blue dot) overlap and interfere constructively in the aperture. The dip field estimated along the structure and the displacement field between the two images are perfectly aligned (b). Horizontal reflector for incorrect velocity model (c); the two images no longer perfectly overlap. The dip and displacement fields form a non-zero angle that measures the inconsistency between the two results (d).

3.1 Objective function from image warping

To measure velocity errors in the experiment-image domain, we need to design the objective function to be optimized. The objective function measures information about the model parameters supplied by the seismic experiments; this information is encoded in terms of, for example, focusing or semblance of prestack images (Symes, 2008). Given two images, we construct two vector fields: the structural dip field and the relative displacement, or warping, vector field. The first can be obtained from either of the considered images; the second links the two images and is computed considering the spatial local crosscorrelation of the two images at every point. If the velocity model is correct, images from adjacent shots feature the same structural dip (this is another expression of constructive interference), and the dip and warping fields are orthogonal (Figure 2(a) and 2(b)). In contrasts, if the velocity model is incorrect, the two vector fields are not pointwise orthogonal (Figure 2(c) and 2(d)); the stack of different experiments is not completely constructive normal to the structural dip and the wavelet in the final image loses its symmetry (in case the data are otherwise zero-phase) (de Vries and Berkhout, 1984). The angle between the dip and warping field therefore measures the quality of constructive interference between the two shots and supplies an error indicator for MVA.

The orthogonality between two vectors is measured by their inner product. The dip field can be estimated using, for example, plane-wave destruction (PWD) filters (Fomel, 2002) or gradient square tensors (van Vliet and Verbeeck, 1995; Hale, 2007a) whereas the warping field is computed using spatial local crosscorrelation of the input images at every spatial location (Hale, 2007b). Figure 3 shows the dip (black arrows) and displacement vector field (white arrows) computed from the migrated images of a horizontal reflector; the velocity error is constant across the model. Note that only when the velocity
Figure 3. Horizontal density interface imaged using a single shot at $x = 2$ km and 800 receivers spaced 10 m at every grid point at the surface. The angle between the dip (white arrows) and displacement vector field (black arrows) indicates a velocity error: (a) velocity too low, (b) correct velocity, and (c) velocity too high. The displacement vector field is computed using a second image obtained from a shot located at $x = 2.05$ km.
model is correct, are the two vector fields orthogonal at every image point. In order to obtain the image and the vector fields, we migrated two shots located at \( x = 2 \) km and \( x = 2.05 \) km, the horizontal interface is due to a density contrast and the velocity model is constant. Our semblance measure can be directly extended to 3D, where the dip vector is normal to the plane tangent to the reflector at every spatial location.

The objective function we define is

\[
J_{vec}(s) = \frac{1}{2} ||d(x) \cdot u(x)||^2, \tag{1}
\]

where \( d(x) = D[x](R_i) \) represents the dip vector field, \( u(x) = U[x](R_i, R_j) \) is the warping vector field between the two images, \( R_i \) and \( R_j \), and the inner product is computed at every image point. The operators \( D[x] \) and \( U[x] \) are the dip and the warping estimation operator, respectively. If the model is correct, the inner products are zero and the objective function \( J_{vec}(s) \) is minimum. The two images \( R_i \) and \( R_j \) are functions of three variables themselves: the experiment index (which we denote by \( i \) and \( j \)), the position vector \( x \), and the model \( s = s(x) \) we used for computing the image:

\[
R_i = R_i(x, s), \quad R_j = R_j(x, s). \]

We parametrize the problem using the slowness rather than velocity because the WEMVA operator is obtained by linearizing the one-way migration operator with respect to slowness (Sava and Biondi, 2004). Note that by measuring the angle between the dip and warping vector fields, we obtain an objective function (equation 1) that is insensitive to amplitude variation as a function of experiment. This feature is particularly helpful because amplitude is a second-order effect with respect to velocity. The warping vector field that maps one image into the other absorbs the amplitude pattern information and makes the approach robust against differences in illumination, geometrical spreading, and reflection coefficient.

In the next section, we show that the objective function obtained from the warping and the vector field is equivalent to the formulation of differential semblance in the experiment-image domain when we consider pairs of adjacent experiments in the limit of small separation between them (e.g., adjacent shots or plane waves with similar ray-parameter).

### 3.2 Image difference and image warping

In this section, we link the objective function in reqobjFunVec to the energy of the image difference using the warping relationship that locally relates two images. The difference operator is a linear operator, and the established link gives us the handle to construct a WEMVA operator and set up the velocity analysis problem as a sequence of linearized inverse problems.

The gradient operator \( \nabla = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right)^T \) applied to the image returns the vector that points toward the maximum increase. It operates as an edge detector and supplies the direction normal to the reflector. Because of the oscillating nature of the signal, the orientation of the gradient vector changes by 180 degrees across the wavelet. Nonetheless, since we are interested in the orthogonality between the dip and warping vector fields, we typically consider the gradient of the image instead of the dip field without loss of information. Hence, we can rewrite equation 1 as

\[
J_{vec}(s) = \frac{1}{2} ||\nabla R_i(x, s) \cdot u(x)||^2. \tag{2}
\]

The warping field between two neighboring images is estimated assuming that locally

\[
R_j(x, s) = R_i(x + u(x), s). \tag{3}
\]

If we also assume that the warping vector \( u(x) \) is small in that the two experiments illuminate the same portion of the subsurface, we can rewrite equation 3 using a Taylor series expansion as

\[
R_j(x, s) \approx R_i(x, s) + \nabla R_i(x, s) \cdot u(x). \tag{4}
\]

Equation 4 allows us to approximate the image difference \( R_j(x, \nu) - R_i(x, \nu) \) as

\[
R_j(x, s) - R_i(x, s) \approx \nabla R_i(x, s) \cdot u(x). \tag{5}
\]

We can now invoke the semblance principle: if the model is correct, the image difference for neighboring experiments must be minimized. Considering equation 5 and computing the energy of \( R_j(x, s) - R_i(x, s) \), we obtain

\[
J_{diff}(s) = \frac{1}{2} ||R_j(x, s) - R_i(x, s)||^2 \approx \frac{1}{2} ||\nabla R_i(x, s) \cdot u(x)||^2, \tag{6}
\]

which is exactly equation 2.

These results hold as long as the two experiments actually illuminate the subsurface in the same aperture. The inner product between the dip and warping field relies on the assumption of specular reflectors, whose dip varies slowly across the image; the constructive interference between different experiments within the aperture ensures that the minimum of the objective function is zero for all practical purposes. We use equation 5 to construct the wave-equation migration velocity analysis operator (Biondi and Sava, 1999; Sava and Biondi, 2004) for our objective function (equation 1).

### 3.3 Wave-Equation Migration Velocity Analysis

In this section, we derive the expression for the wave-equation migration velocity analysis (WEMVA) operator (Biondi and Sava, 1999; Sava and Biondi, 2004) for our objective function (equation 1). A detailed presentation of the WEMVA procedure can be found in Sava and Biondi (2004); implementation aspects are described by Sava and Vlad (2008).

The WEMVA operator is based on a linearization of the extrapolator operator and the single-scattering (Born) approximation. The main goal of this method is to link a perturbation \( \Delta s \) in the slowness field with the perturbation \( \Delta R \) in the image that is obtained after migration. Because of the linearization, the link is described by a linear operator.

The image \( R(s) \) depends on the slowness model \( s \) through the source and receiver wavefields \( W_S(s, \omega) \) and
\[ W_R(s, \omega) \]
\[ R(s) = \sum_\omega W_S(s, \omega) W_R(s, \omega), \]
where the overline indicates the complex conjugate operator and \( \omega \) represents the temporal frequency. The wavefields are constructed in the frequency domain using a one-way phase-shift operator (Sava and Vlad, 2008). In equation 7, the image \( R(s) \) depends nonlinearly on the model \( s \); the nonlinearity comes from the wave-equation, which we have to solve for computing the source and receiver wavefields \( W_S(s, \omega) \) and \( W_R(s, \omega) \).

To linearize the relationship between the image and model, we assume a “small” perturbation \( \Delta s \) that would produce small perturbations in the wavefields. In the following, we drop the dependence on \( \omega \) and \( s \) for the sake of readability.

By perturbing equation 7 with respect to the slowness field, we obtain
\[ R + \Delta R \approx \sum_\omega \left( \overline{W_S} W_R + \Delta W_S W_R + \overline{W_S} \Delta W_R \right), \]
where \( \Delta W_S \) and \( \Delta W_R \) are the linear perturbations in the source and receiver wavefields resulting from a perturbation \( \Delta s \) in the slowness model. \( \Delta R \) represents the perturbation that would be observed in the image.

Using equation 8, we can relate the image perturbation \( \Delta R \) and the wavefield perturbation by removing the background image from the left and right side of the equation to obtain
\[ \Delta R \approx \sum_\omega \left( \Delta W_S W_R + \overline{W_S} \Delta W_R \right) \approx L \Delta s. \]
\[ (9) \]

In equation 9, the wavefield perturbations \( \Delta W_S = \Delta W_S(\Delta s) \) and \( \Delta W_R = \Delta W_R(\Delta s) \) depend linearly (Born approximation) on the slowness perturbation \( \Delta s \) (Sava and Vlad, 2008). Thus the image perturbation \( \Delta R \) is obtained by applying a linear operator \( L \) to the model perturbation \( \Delta s \) (Sava and Biondi, 2004).

This result is the basis for the derivation of the WEMVA operator for our objective function. Let us consider two shots and build the image perturbation for the individual shots and their difference for a certain model perturbation \( \Delta s \):
\[ \Delta R_i = L_i \Delta s \]
\[ \Delta R_j = L_j \Delta s \]
\[ \Delta R_j - \Delta R_i = (L_j - L_i) \Delta s; \]
(10)
the operators \( L_i \) and \( L_j \) differ because they are constructed using wavefields from the two experiments \( i \) and \( j \), even though they use the same background model. Equation 10 shows that we can linearly combine the WEMVA operator to obtain the image perturbation for a linear combination of a set of images. This result follows from the linearity of the operator itself. Before assuming that equation 10 is the definition of the WEMVA operator for our objective function, we need to show that the right-hand side does represent an image perturbation. Let us consider two images \( R_i = R_i^c + \Delta R_i \) and \( R_j = R_j^c + \Delta R_j \). Each of them is the superposition of the image in the background correct model \( R_i^c \) and \( R_j^c \) and the image perturbation caused by the model perturbation \( \Delta s \) \( (\Delta R_i \) and \( \Delta R_j \) ); then the difference of the two images is
\[ R_j(x) - R_i(x) = R_j^c(x) - R_i^c(x) + \Delta R_j - \Delta R_i. \]
(11)
If the shots are close enough, the similarity of the images obtained in the correct background model is maximum, and their difference in equation 11 goes to zero:
\[ R_j(x) - R_i(x) \approx \Delta R_j - \Delta R_i. \]
(12)
Equation 12 shows that under the assumption of small separation between the shots, the image difference \( R_j(x) - R_i(x) \) gives us an approximation of the image perturbation we need for the WEMVA operator.

3.4 Image perturbation

The formulation of the optimization problem for velocity analysis discussed in the preceding sections gives a handle for constructing an image perturbation that drives the velocity update in wave-equation migration velocity analysis (WEMVA) (Biondi and Sava, 1999; Sava and Biondi, 2004). As previously discussed, an objective function based on the image difference is potentially sensitive to differences in amplitudes, for example because of the presence of strong shallow lenses. A model update attributable only to amplitude differences is a serious drawback that we want to avoid.

Again, the objective function based on the dip (gradient) and warping vector fields is purely kinematic: the measure of orthogonality between the two fields, i.e., the measure of constructive interference between the data for the two shots, exploits amplitudes in the images but it is not biased by them. Moreover, it automatically leads to an image perturbation \( \Delta R \) that is asymptotically equivalent to the image difference. Consider again the two images \( R_i \) and \( R_j \) and assume that they are related by the expression in equation 3. By linearizing equation 3, we obtained the relationship between the difference of the images and the product of the gradient of the image and the displacement field. We can thus consider the first-order term of equation 4,
\[ \Delta R = \nabla R_i(x, s) \cdot u(x), \]
(13)
as the image perturbation for WEMVA. Since the gradient of the image is parallel to the dip vector when the model is correct, the image perturbation \( \Delta R \) in equation 13 is zero for all practical purposes (whereas the image difference is not), and we avoid a velocity update driven only by differences in the amplitudes between the two images.

4 NUMERICAL EXAMPLES

We next present a few numerical examples to describe our methodology. First, we show the behavior of the objective function for a simple model. Second, we show the computed
Figure 4. Vertical section of a 3D model with azimuthal symmetry. (a) Density model, (b) background slowness model, and (c) slowness anomaly.

image perturbations and backprojections using the adjoint WEMVA operator. The first inversion test considers a simple model with a stack of horizontal density interfaces in a vertical velocity gradient; we perturb the slowness model with two Gaussian anomalies with opposite sign, run 10 iterations of nonlinear conjugate gradient, and correct the model. We conclude the section showing the inversion results obtained on the marmousi dataset using a very limited number of shot-gathers. Our methodology is able to correct the model and obtain a more geologically plausible image.
4.1 Objective functions

Consider a horizontal reflector in a 3D model with a vertical slowness gradient (Figure 4(a) and 4(b)). The interface is a density contrast, and the data are modeled with a two-way finite-difference algorithm. We contaminate the data with Gaussian random noise; the signal-to-noise ratio for each shot-gather is fixed and equal to 10. We first consider a 2D slice of the model and run a 2D sensitivity test. We migrate the data with a downward continuation operator (Stoffa et al., 1990) and different amplitudes of the velocity anomaly in Figure 4(c). We migrated two shots located at $x = 2$ km and $x = 2.05$ km changing the value of the anomaly in the model to perform a sensitivity analysis. Figure 5 shows the computed objective functions from the image difference (5(a)) and image warping approach (5(b)). Both measures lead to smooth and monomodal objective functions in a wide range of realistic velocity errors (from $-0.4$ km/s to $0.5$ km/s peak amplitude). The similarity between the behavior of the two objective functions supports our claim that the two approaches are actually connected for finite shot separation, as shown in the previous sections. Nonetheless, the minimum value of image difference objective function is not zero (Figure 5(a)) because of the different amplitude pattern in the two images. In the 3D case with the azimuthally symmetric anomaly in Figure 4(c), our image warping approach produces a distinctly steeper objective function (Figure 6), which implies more sensitivity to velocity errors. The image difference approach (6(a)) produces a strong nonzero value for the correct model in the objective function compared with the warping approach (Figure 6(b)). Moreover, the objective function for the image difference is much less sensitive to velocity errors. We observed that this problem is mainly due to the healing of the wavefield below the anomaly that produces a fairly continuous reflector even for incorrect models. The healing of the wavefield is more prominent in 3D than in 2D.
Figure 7. Image perturbations for a negative slowness anomaly computed from (a) the exact image perturbation obtained from the forward WEMVA operator and (b) image warping. Associated backprojections computed from (c) the exact image perturbation obtained from the forward WEMVA operator and (d) image warping. The phase of the wavelet of the estimated image perturbation is consistent with the ideal case, and the sign of the backprojection matches the sign of the anomaly.

Figure 8. Image perturbations for a positive slowness anomaly computed from (a) the exact image perturbation obtained from the forward WEMVA operator and (b) image warping. Associated backprojections computed from (c) the exact image perturbation obtained from the forward WEMVA operator and (d) image warping. The phase of the wavelet of the estimated image perturbation is consistent with the ideal case, and the sign of the backprojection matches the sign of the anomaly.
4.2 Image perturbations and backprojections

The information about the model error is mainly in the phase of the image perturbation. To assess the quality of the image perturbation we obtain from the migrated images, we compare the ideal image perturbation, obtained from the direct WEMVA operator for a given slowness perturbation $\Delta s$, with the image perturbation obtained from the warping vector field. The image perturbation represents the data of a linear system of equations that link the model perturbation $\Delta s$ to the image perturbation $\Delta R$ via a linear operator $L$. The adjoint of the operator $L$ maps the input data back in the model space and gives a scaled version of the model perturbation $\Delta s$.

Figure 7 and Figure 8 show the image perturbations and the associated backprojections computed directly from the WEMVA operator and using our image warping approach; we show two cases using a single horizontal reflector and gaussian slowness anomalies of different sign. We want to validate our procedure showing that the image perturbations and the associated backprojections we compute are consistent with the ideal case.

Figure 7(a) and 7(c) show the image perturbation that a slowness perturbation produces and the relative backprojection, which is obtained by applying the adjoint WEMVA operator to the image perturbation. They represent the ideal result since in real scenarios we do not have access to the model perturbation (which we want to estimate). In a realistic situation, we estimate the image perturbation and backproject it to estimate the slowness anomaly responsible for it. Figure 7(b) shows the estimate obtained using the image warping relation between neighboring images; by applying the adjoint of the WEMVA operator to this image, we obtain the backprojection in Figure 7(d). Observe the consistency with the backprojection in Figure 7(c). Figure 8 shows the case for a model anomaly with opposite sign. Observe that the backprojection changes sign accordingly. Note that our estimate of the image perturbation leads to a backprojection that is consistent with the ideal case scenario in which we actually know the model perturbation.

The backprojection of the estimated image perturbation contains sidelobes, which have been observed in other migration velocity analysis algorithms (Fei and Williamson, 2010); we conjecture that these artifacts are attributable to the actual nonlinearity of the estimated image perturbation in terms of the model parameters (slowness in this case).

4.3 Inversion test

As mentioned before, the estimation of model parameters for wave phenomena is a nonlinear problem because the wave-equation is nonlinear in the model parameters (slowness, density, etc.). In order to test our inversion methodology, we use a standard implementation of the nonlinear conjugate gradient algorithm (Vogel, 2002), in which the nonlinear problem is first linearized and an approximate solution is obtained. The approximate solution is then used to obtain a new linearization and then a new approximate solution, and so on. For every nonlinear iteration, we recompute the images and solve a linearized problem assuming the new background medium. The model for the inversion is shown in Figure 9: the reflectors arise from density constrasts (Figure 9(a)) and the correct slowness model is a simple vertical gradient (Figure 9(b)). We consider 24 pairs of shots at the surface; the receiver are evenly spaced by 10 m at each grid point at the surface. The shots in each pair are 50 apart and the spacing between the pairs is 250 meters; the shot pairs do not overlap.

The initial and updated velocity model are shown in Figure 10. The data for the inversion test are computed through full-acoustic (i.e., non-Born) time-domain finite-difference modeling and Gaussian random noise is added to each shot-gather separately. The signal-to-noise ratio is equal to 10. Absorbing boundary conditions are implemented to avoid surface-related multiples; the internal multiples generated by the reflectors in the model are not subtracted from the data and constitute additional source of noise.

We run 10 nonlinear iterations involving five conjugate-gradient iterations for each linearized step. The Gaussian shape anomalies in Figure 10(a) are smoothed out by the migration velocity analysis procedure (Figure 10(b)). Figure 11 shows the resulting migrated images for the initial velocity model (Figure 11(a)) and that after five iterations of the nonlinear conjugate gradient (Figure 11(b)). Observe the improved focusing of the image, especially for the deepest reflectors.

In the examples, the absolute value of the peak of the velocity anomaly is 200 m/s, which is 10% of the minimum value of the background velocity gradient. Nonetheless, the shape of the velocity anomaly causes severe focusing/defocusing of the wavefield (depending on the sign of the anomaly), which makes migration velocity analysis challenging, especially in the initial steps. Figure 12 shows the decrease of the objective function with the number of iterations. The nonzero value of the objective function results from a number of factors: the random noise in the data, internal multiples, and the size of the local window used to compute the dip and displacement field; all of them constrain the accuracy, resolution, and sensitivity we can achieve.

4.4 Marmousi

In this section, we consider a more complex synthetic example. Figure 13 shows the correct macro velocity model for the Marmousi dataset and the anomaly used in the inversion test. We want to show the ability of our method of recovering model perturbation using a limited number of experiments (shots in this case); we consider 12 shots, evenly spaced by 0.04 km at the surface; the first shot is positioned at $x = 0.96$ km. The receivers are spaced 0.008 km on the surface and they are located at every grid point.

We perturb the model in Figure 13(a) with the anomaly in Figure 13(b); the maximum amplitude of the anomaly is 0.032 s/km, which is about 10% of the local background slowness. Figure 14(a) shows the image of the 12 shot-gathers considered in the correct model; by comparing the correct image with the image obtained using the initial model (Figure 14(b)) we observe that several structural features are distorted and
the continuity of some reflectors broken (compare reflectors at lateral position \( x = 2 \text{ km} \)). After 30 iterations of nonlinear conjugate-gradient we obtain the image in Figure 14(c): the bottom reflector is now more continuous and both the shape and amplitude of the dipping reflector in the middle section have been corrected; we want to highlight that, overall, we improved the image quality using a very limited number of shotgathers. We also want to stress that these results do not involve any regularization or shaping of the estimated anomaly, which would greatly speed up convergence. Figure 15 shows the evolution of the objective function (the energy of the image perturbations) with the number of iterations; the objective function rapidly decreases to 50% of the initial value and then flattens. Since we use only a limited number of experiments (12), we introduce migration noise in the image, which prevents the inversion algorithm from further reducing the objective function. We want to emphasize, however, that we are able to correct the anomaly in the model using a small number of shots despite the lack of full illumination whereas methods based on focusing measures (i.e. migration velocity analysis based on extended images) need complete angular coverage to evaluate focusing in the image domain.

5 DISCUSSION

Our objective function measures the constructive interference of pairs of images using the warping field that relates them and the structural dip. The velocity error produces shifts of the image points, which can be observed in the image cube as moveout as a function of an extension parameter (e.g., the reflection angle or the shot index). The moveout represents a velocity error indicator and is inverted for updating the velocity model. In the literature (Biondi and Symes, 2004a; Sava and Fomel, 2006; Xie and Yang, 2008), the assumption of a shift along the normal to the reflector is commonly made; however,
as Xie and Yang (2008) point out, this is just an approximation in a general case of complex geology and velocity model. Moreover, the residual moveout is often measured along just the vertical direction, i.e., in CIGs, rather than along the normal to the reflector, and then converted into the normal residual moveout using simple trigonometry. That approach relies on the assumption that the dip of the reflector is independent of the velocity in the overburden. In general, this is not true. Our approach directly evaluates the degree of constructive interference of adjacent experiments and does not rely on any assumption about the direction of the shift of the image point in the migrated domain. Moreover, we are able to measure the 3D shift of the image point between the two experiments and then capture the complete kinematic information.

Operating in the image space, we can analyze all points in the portion of the model illuminated by the two experiments simultaneously and recover a spatially extended indicator of velocity error within the aperture of the experiments. This approach is computationally convenient because the construction of CIGs requires the subsampling of the image space for handling the dimensionality of the problem and the associated computational cost, whereas our method does not require gathers and supplies reliable information about the velocity inconsistencies in the area of the model illuminated by the seismic experiments.

Migration velocity analysis is based on kinematic assumptions: the correct velocity model returns a prestack image cube characterized by horizontal events along the exper-
Figure 11. Migrated images: (a) image obtained with the initial velocity model and (b) after 10 velocity updates. The image in (b) is obtained after 10 nonlinear conjugate gradient iterations; every linearized problem involves five iterations of conjugate gradient. Note the better focusing of the image.

iment axis. Note that we say horizontal and not constant in amplitude, highlighting the kinematic nature of the conventional semblance principle. As already discussed by Mulder and Kroode (2002), the amplitudes of images from nearby experiments are likely to be similar but not equal. Reflection coefficients should be taken into account, and the velocity model itself, if it is highly heterogeneous, can lead to important differences in the amplitudes.

The image-domain implementation and the differential semblance approach makes our strategy (as opposed to WI) robust against local minima and cycle skipping; the measure of velocity error is independent of the migration algorithm and in principle either a one-way or a two-way velocity update engine can be used. The implementation of the WEMVA procedure, in contrast, relies on a one-way operator and cannot be directly extended to two-way operators.

The differential semblance approach makes both strategies discussed in this paper robust against cycle skipping. If the current velocity model is reasonably close to the true model, a cycle skip of the two images is unlikely. The chance of cycle skipping depends not only on the velocity model but also on other factors such as the shot positions and the frequency bandwidth of the signals. By increasing the distance between the shots, we violate the assumption of sharing the same aperture and the potential for cycle skipping grows. Increasing the bandwidth of the signals reduces the wavelength,
Figure 12. The decrease of objective function indicates that we are approaching the correct model.

Figure 13. Macro slowness model (a) for the Marmousi dataset and (b) slowness perturbation for the model in (a).
Figure 14. Image obtained using the correct slowness model (a), initial migrated image (b), migrated image after 30 iterations of nonlinear conjugate-gradient (c). We consider 12 shots, evenly spaced by 0.04 km at the surface; the first shot is positioned at $x = 0.96$ km. The receivers are spaced 0.008 km on the surface and they are located at every grid point. Observe the improved continuity of the bottom reflector and the correction of the distortion of the middle dipping interfaces at a lateral position about $x = 2$ km.
thus further constraining the maximum distance between adjacent experiments. For the image difference, the shot distance enhances the amplitude mismatch, thus the image perturbation can experience both cycle skipping and incomplete event cancellation, especially for strong and shallow reflectors. The construction of the image perturbation as the first order term of a Taylor series expansion removes the sensitivity to amplitude mismatching. Also cycle skipping is avoided since we are not computing differences but we construct the image perturbation from a single image and the estimated warping vector field.

Several critical points must be carefully handled. First, since we are using only pairs of experiments, the signal-to-noise ratio (SNR) can be low and the effectiveness of the algorithm may be hampered by the quality of the data. The sensitivity to the signal-to-noise ratio can be addressed using shot-encoding techniques (Whitmore, 1995; Zhang et al., 2005; Morton and Ober, 1998; Soubaras, 2006; Perrone and Sava, 2011) under the assumption that the different experiments are contaminated by uncorrelated noise. Second, because of cycle skipping between images, our approach can be ineffective if the shots are too far from each other; the velocity model is very inaccurate, or the reflector geometry is complex (e.g. around fault zones or in areas with conflicting dips). The similarity of nearby experiments is an assumption of every DSO-like optimization scheme; if the particular acquisition geometry or the complex wavepaths create shadow zones or illumination holes, DSO is likely to underperform. In the shot-domain, diving waves create images that do not correspond to reflection events and that do not satisfy the warping relationship assumed in this paper. Analogous considerations apply to multiply scattered wave and converted waves, which are not correctly handled by migration algorithms based on the Born approximation. Apparent reflectors associated with multiples and converted waves do not satisfy the warping relationship we assumed for the definition of the objective function and thus introduce a systematic bias in the inversion scheme.

6 CONCLUSIONS

We introduce a new measure of velocity errors for migration velocity analysis in the shot-domain. Our measure is based on the structural information in the image (the dip vector field) and a second vector field that links the images of adjacent experiments. This second vector field warps the first image into the second and captures the apparent displacement attributable to the shift in source position (in conventional shot-gather migration) or different ray-parameter (in plane-wave migration). The correct velocity model leads to images that build up along the reflectors; hence, by measuring the angle between the dip and the warping vector fields we can estimate velocity errors. Our criterion restates the semblance principle and takes into account the motion of the image points as a function of shot position and velocity model, instead of considering gathers and moveout in the vertical direction only. The method allows one to assess the correctness of the velocity model at every point in the image space within the aperture of the two considered shots; moreover, since no gather is required, one can bootstrap the velocity-analysis step while imaging. Possible downsides are related to reduced the signal-to-noise ratio, which is smaller because we consider groups of experiments at a time instead of the entire dataset. The formulation of the optimization problem for the velocity model naturally leads to an expression for an image perturbation for wave-equation migration velocity analysis. The image perturbation is computed as the inner product between the image gradient and the warping vector field at every image point, i.e., the first-order term of a pointwise Taylor series expansion based on the warping relationship between the two images. If the velocity model is correct, the image gradient (orthogonal to the structure) and the displacement field (parallel to the structure) are orthogonal, the objective function is minimum, and the image perturbation is zero. A synthetic test involving local velocity anomalies and severe focusing and defocusing of the wavefields show the effectiveness of our approach for linearized migration velocity analysis.
ACKNOWLEDGMENT

We would like to acknowledge the financial support provided by a research grant from eni E&P. The reproducible numeric examples in this paper use the Madagascar open-source software package (http://www.reproducibility.org). We thank Jeff Godwin for integrating the Mines JTK (http://inside.mines.edu/~dhale/jtk/) and the Madagascar.

REFERENCES

Vogel, C., 2002, Computational methods for inverse problems: SIAM.
Yilmaz, O., 2001, Seismic data processing: SEG.
Wavefield tomography based on local image correlations

Francesco Perrone\(^1\) and Paul Sava\(^1\)
\(^1\) Center for Wave Phenomena, Colorado School of Mines

ABSTRACT

The estimation of a velocity model from seismic data is a crucial step for obtaining a high-quality image of the subsurface. Velocity estimation is usually formulated as an optimization problem where an objective function measures the mismatch between synthetic and recorded wavefields and its gradient is used to update the model. The objective function can be defined in the data-space (as in full-waveform inversion) or in the image-space (as in migration velocity analysis). In general, the latter leads to smooth objective functions, which are monomodal in a wider basin about the global minimum compared to the objective functions defined in the data space. Nonetheless, migration velocity analysis requires construction of common-image gathers at fixed spatial locations and subsampling of the image in order to assess the consistency between the trial velocity model and the observed data. We present an objective function that extracts the velocity error information directly in the image domain without computing common-image gathers. Because of the dimensionality of the problem, gradient-based methods (such as the conjugate-gradient algorithm) are used in the optimization procedure. In order to include the full complexity of the wavefield in the velocity estimation algorithm, we consider a two-way (as opposed to one-way) wave operator, we do not linearize the imaging operator with respect to the model parameters (as in linearized wave-equation migration velocity analysis), and compute the gradient of the objective function using the adjoint-state method. We illustrate our velocity estimation methodology with a few synthetic examples.

Key words: tomography, migration velocity analysis, imaging

1 INTRODUCTION

Seismic imaging involves the estimation of wave propagation velocities in the subsurface from seismic data recorded at the surface. Seismic velocities are related to other physical parameters (for example, density and compressibility, which characterize the lithology of the Earth), and rock-mechanics parameters (for example, porosity and fluid overpressure, which are crucial in reservoir engineering) (Carcione, 2007).

Seismic imaging includes the estimation of both the position of the structures that generate the data recorded at the surface and of a model that describes wave propagation in the subsurface. The two problems are closely related since a model is necessary to infer the position of the reflectors. Waves recorded at the surface are extrapolated in a model of the subsurface (by solving a wave equation) and crosscorrelated with a synthetic source wavefield simulated in the same model (Claerbout, 1985). Under the single scattering approximation, reflectors are located where the source and receiver wavefields match in time and space.

Wave-equation tomography (Tarantola, 1984; Woodward, 1992; Biondi and Sava, 1999) is a family of techniques that estimate the velocity model parameters from finite bandwidth signals recorded at the surface. The inversion is usually formulated as an optimization problem where the correct velocity model minimizes an objective function that measures the inconsistency between a trial model and the observations. The objective function can be defined either in the data-space (full-waveform inversion) or in the image-space (migration velocity analysis).

Full-waveform inversion (FWI) (Tarantola, 1984; Pratt, 1999; Sirgue and Pratt, 2004) addresses the estimation problem in the data-space and measures the mismatch between the observations and simulated data. Full-waveform inversion aims to reconstruct the exact model that generates the recorded data. By matching both traveltimes and amplitudes, full-waveform inversion allows one to achieve high-resolution (Sirgue et al., 2010). Nonetheless, a source estimate is needed, the physics of wave propagation (for example, isotropic vs. anisotropic, acoustic vs. elastic, etc.) must be correctly modelled, and a good parametrization (for example, impedance vs. velocity contrasts) is crucial (Kelly et al., 2010). Moreover, an accurate initial model is key to avoid cycle skipping and con-
verge to the global minimum (instead of a local minimum) of the objective function.

Because of the nonlinearity of the wavefields with respect to the velocity model, the objective function in the data domain is highly multimodal (Santosa and Symes, 1989), and local optimization methods can easily get trapped into local minima and fail to converge to the correct model. This is particularly true for reflection full-waveform inversion. Refraction full-waveform inversion focuses on diverging waves and mutes the data, retaining only the diving energy (Pratt, 1999). This leads to a better-behaved objective function but requires very long offsets in order to record the refracted energy. Moreover, this approach limits the depth at which a robust inversion result can be expected.

Migration velocity analysis (MVA) (Fowler, 1985; Faye and Jeannot, 1986; Al-Yahya, 1989; Chavent and Jacewitz, 1995; Biondi and Sava, 1999; Sava et al., 2005; Albertin et al., 2006) defines the objective function in the image space and is based on the semblance principle (Al-Yahya, 1989). If the velocity model is correct, images from different experiments must be consistent with each other because a single Earth model generates the recorded data. A measure of consistency is usually computed through conventional semblance (Taner and Koehler, 1969) or differential semblance (Symes and Carazzone, 1991). These two functionals analyze a set of migrated images at fixed locations in space; they consider all the shots that illuminate the points under investigation. Migration velocity analysis leads to smooth objective functions and well-behaved optimization problems (Symes, 1991; Symes and Carazzone, 1991) and is less sensitive than full-waveform inversion to the initial model. On the other hand, because we do not use amplitudes in the imaging step, the estimated model has lower resolution than the ideal full-waveform inversion result (Uwe Albertin, personal communication).

Migration velocity analysis measures either the invariance of the migrated images in an auxiliary dimension (reflection angle, shot, etc.) (Al-Yahya, 1989; Rickett and Sava, 2002; Sava and Fomel, 2003; Xie and Yang, 2008) or focusing in an extended space (Rickett and Sava, 2002; Symes, 2008; Sava and Vasconcelos, 2009; Yang and Sava, 2011). All these approaches require the migration of the entire survey in order to analyze the moveout curve in common-image gathers or measure focusing at a specific spatial location. The dimensionality of the (extended) image space and computational complexity of the velocity analysis step rapidly explodes for realistic case scenarios. Moreover, because of the high memory requirement for storing the partial information from each experiment, only a subset of the image points can be considered in the evaluation of the objective function. Illumination holes and/or irregular acquisition geometries can also impact the quality of the common-image gathers but no systematic study of this problem is reported in the literature to our knowledge.

Reverse-time migration (Baysal et al., 1983; McMechan, 1983) is routinely used in exploration geophysics because of its ability to correctly handle the full complexity of the wave propagation phenomena (under the assumption that the physics of wave propagation in the subsurface is correctly modeled). Its computational cost is nevertheless still prohibitive and limits its integration into a migration velocity analysis loop that requires the extraction of common-image gathers for moveout analysis. A migration velocity analysis procedure based on a full-wave propagation engine allows exploitation of more complex wave phenomena (e.g., overturning reflection, prismatic waves, and multiples) and increases the amount of information in the data that can be used (Farmer et al., 2006). Nonetheless, because of the intrinsic cost of wave extrapolation in the time domain, a different approach must be considered for an effective and efficient implementation of velocity model building.

Up to now, the powerful tool offered by reverse-time migration has been used only for reconstructing an image of the impedance contrasts in the subsurface. The velocity model used for generating a migrated image is obtained with other techniques, for example through traveltime tomography (Bishop et al., 1985) or through wave-equation migration velocity analysis (Biondi and Sava, 1999; Sava and Biondi, 2004; Shen and Symes, 2008), which are based on asymptotic approximations and/or linearization of the wave equation operator. These techniques do not exploit all the information encoded in the recorded wavefields because they are either inaccurate in complex velocity models (traveltime tomography) or unable to properly model part of the propagation phenomena (one-way methods). Designing a velocity model-building procedure based on a two-way wave propagation engine will allow us to exploit the full complexity of the wavefields and will make the velocity analysis step conceptually consistent with the imaging algorithm used.

Because reverse-time migration is so demanding from a computational and storage point of view, we propose to analyze the semblance of small groups of migrated images from nearby experiments, i.e. neighboring shots or plane-waves with similar ray-parameter or take-off angle, at every image point illuminated by the experiments. Reverse-time migration works only in common-shot and common-receiver configurations, so the choice of the shot-domain for performing migration velocity analysis with a two-way engine may be preferable. Here, we use the word “shot” in its broader sense to include synthetic shot-gathers like plane-wave sources (Whitmore, 1995; Liu et al., 2006; Stoffa et al., 2006; Zhang et al., 2005), random shot-encoded sources (Morton and Ober, 1998; Romero et al., 2000) or any other phase/amplitude-encoded source (Soubaras, 2006; Perrone and Sava, 2011) in the range of shot-profile migration. Other techniques for velocity model building in the shot-domain use different strategies. Nonetheless, all of them consider the entire survey and migrate all the data before starting the migration velocity analysis loop (de Vries and Berkhourt, 1984; Yilmaz and Chambers, 1984; Al-Yahya, 1989; Chavent and Jacewitz, 1995; Sava and Biondi, 2004; Shen and Symes, 2008; Symes, 2008; Xie and Yang, 2008). We propose an objective function that evaluates the degree of semblance between images through local correlations in the image space and does not need common-image gathers (CIGs). We use the morphologic relationship between
images from nearby experiments to define an objective function that measures shifts in the image space. The methodology described in this work follows from the linearized wave-equation MVA operator proposed by (Perrone et al., 2012) but removes the linearization of the wave-operator with respect to the model parameters and thus accounts for the full two-way nature of the wavefields. We compute the gradient of the objective function with the adjoint-state method (Plessix, 2006). Our approach reduces the memory requirements and avoids the need to pick moveout on gathers, and allows us to include all points illuminated by the seismic experiments in the velocity model building.

2 THEORY

In order to set up an optimization problem, we need to define a residual in either the data space, \( r(t) \), or the image space, \( r(x) \), that represents an indirect measure of the error in the model parameters. An objective function can then be defined as the energy (\( l_2 \) norm) of the residual in the appropriate domain. Other norms (for example the \( l_1 \) norm) have been used to exploit the concept of sparsity and gain robustness against the unavoidable noise in the data (Claerbout and Muir, 1973; Tarantola, 2005). In the simple case of linear problems, the \( l_2 \) norm leads to quadratic optimization problems for which we can easily compute the gradient of the objective function; in contrast, the \( l_1 \) norm leads to nonquadratic objective function which are much more difficult to differentiate. This fundamental difference makes the \( l_2 \) norm a more popular choice for setting up an inverse problem in the optimization framework. In the following, we briefly review how the residual is defined; then, we introduce our measure of velocity error in the image space and the resulting objective function. The adjoint-state calculation of the gradient concludes the section.

2.1 Data space residuals

The residual \( r(t) \) can be defined in different ways: in fullwaveform inversion, the residual is a function of time, and it is defined as the difference between the synthetic and observed data (Tarantola, 1984):

\[
r(t) = d^{syn}(t) - d^{obs}(t).\]

This residual captures information about the traveltine and amplitude error due to an incorrect velocity model. The observed data set a time reference and the mismatch in traveltine can be directly backprojected in the velocity model: a positive time lag corresponds to a velocity model that is too fast; a negative time lag corresponds to a velocity model that is too slow. A different residual function can be obtained computing the crosscorrelation of the synthetic and observed data and penalizing the nonzero lags of the crosscorrelation (van Leeuwen and Mulder, 2008):

\[
r(t) = \sum_{\tau} P(\tau) \left( d^{syn}(t) \ast_r d^{obs}(t) \right).\]

The asterisk denotes the crosscorrelation operation, the variable \( \tau \) indicates the crosscorrelation lag in the time domain, the penalty operator \( P(\tau) \) can be any monotonic function of \( \tau \) that is zero at \( \tau = 0 \). In this case, the focus is only on the traveltine information and not on the amplitude error in the data. If the velocity model is incorrect, the crosscorrelation of the synthetic and observed data shifts from the zero lag and indicates the sign of the velocity error. The penalty operator annihilates the signal around zero lag and highlights the traveltine mismatch.

2.2 Image space residuals

In migration velocity analysis, the residual \( r(x) \) is computed using the source and receiver wavefields extrapolated in the subsurface model, and thus it is defined in the image space. Various measure of velocity error are reported in the literature, most of them are based on the curvature of the movout observed in common-image gathers in various domains (offset, reflection angle, etc.). Recently, the focusing criterion (Shen and Symes, 2008; Symes, 2008; Sava and Vasconcelos, 2009) have been used for defining a residual in the image domain that represent a proxy for the velocity error. Both the construction of common-image gathers and the measure of focusing require the migration of all the data that illuminate the portion of the model under investigation.

We present an objective function that does not require the explicit construction of common-image gathers and operates directly on the migrated images. The design of this objective function follows from previous work on linearized wave-equation MVA (Perrone et al., 2012), which used the warping relationship between images to construct an image perturbation that measures the model error and drives the model update. Here, we estimate the relative shifts between images using penalized local correlations (a bilinear operator) and are able to integrate this measure into a full-wavefield inversion algorithm using the adjoint-state method. Our goal is to assess the correctness of the velocity model from a small subset of images from nearby experiments but including all the points in the image into the objective function and without increasing the computational cost. We want to trade complete illumination (necessary to measure focusing) at a single spatial location for subsurface aperture (i.e., all the points in the image illuminated by the shots under analysis). We review the measure of consistency based on the semblance principle and introduced in Perrone et al. (2012), define a residual \( r(x) \) in the image space, and then formulate an objective function as the \( l_2 \) norm of the residual.

There are at least two options for measuring similarity between two images. The first one is to compute the pointwise difference, in which case the residual function is

\[
r(x) = R_{i+1}(x) - R_i(x), \tag{1}\]

where \( R_i(x) \) and \( R_{i+1}(x) \) are the migrated images for \( i \)th and \( i+1 \)th shot, respectively. The two shots are assumed to be close and to image the same area in the subsurface. Plessix (2006) uses the energy of the image difference as a regulariza-
tion term for full-waveform inversion and not as a stand-alone objective function. Notice that if the velocity model used in migration is severely inaccurate, we can have cycle skipping in the image space or, in other words, the difference between the two images can produce two events if the position of the same reflector in the two images changes more than a quarter of a wavelength. This problem is analogous to what happens in full-waveform inversion in the data space.

An alternative solution is to compute local correlations of the two images at each point \( x \) and penalize according to the local dip estimated from one of the images. This approach measures the relative shift of one image with respect to the other directly in the image space and can be used for velocity analysis. Our methodology is the dual in the image domain of the data correlation strategy developed by van Leeuwen and Mulder (2008) in the framework of full-waveform inversion. We express the residual as

\[
  r(x) = \int P(x, \lambda) c_i(x, \lambda) \, d\lambda, \tag{2}
\]

where

\[
  c_i(x, \lambda) = \int_{w(x)} R_i(\xi - \frac{\lambda}{2}) R_{i+1}(\xi + \frac{\lambda}{2}) \, d\xi \tag{3}
\]

is the local correlation of the images \( R_i(x) \) and \( R_{i+1}(x) \) (Hale, 2007b) and \( P(x, \lambda) \) is a penalty operator that highlights the shift of the local correlation along the dip direction. The vector \( \lambda \) denotes the correlation lag in the image space and the index \( i \) scans the shot position. Local Gaussian windows \( w(x) \) centered about the image points \( x \) weight the input images and allow computation of local crosscorrelations.

### 2.3 The imaging condition

The residual \( r(x) \) in equation 2 is computed from the migrated images \( R_i(x) \) and \( R_{i+1}(x) \). The dependency on the source and receiver wavefields is hidden in the imaging condition we use to construct the migrated image. An image of the subsurface \( R(x) \) is conventionally computed as the zero-lag time correlation of the source and receiver wavefields:

\[
  R(x) = \int_T u_s(x, t) u_r(x, t) \, dt, \tag{4}
\]

where \( T \) is the recording time interval in the data, or the zero-lag frequency correlation of the source and receiver wavefields:

\[
  R(x) = \frac{1}{\Omega} \int_\Omega \overline{u_s(x, \omega)} u_r(x, \omega) \, d\omega, \tag{5}
\]

where the overbar denotes the complex conjugation operator and \( \Omega \) denotes the frequency bandwidth of the signals.

### 2.4 Image correlation objective function

We restate the semblance principle as follows: if the velocity model is correct, two images from nearby experiments construct the final image along the direction of the reflectors (Perrone et al., 2012). This is equivalent to the standard assumption that, if the model is correct, the prestack image-cube is invariant with respect to the shot position, i.e., no moveout in the shot-domain common-image gathers (Xie and Yang, 2008).

The structural dip is an attribute that is commonly extracted in seismic processing (van Vliet and Verbeek, 1995; Fomel, 2002; Hale, 2007a). The similarity of two images along the structural dip can be measured by means of local correlations (Hale, 2007b). We illustrate this idea on a simple model with a single horizontal density interface (Figure 1); the correct velocity model is constant and equal to 2.0 km/s. Figure 2 shows the dip and apparent displacement vector field (white and black arrows, respectively) at particular spatial locations. Observe that the two vectors are orthogonal when the velocity model is correct.

The computation of the displacement field requires a highly nonlinear procedure, which extracts the maximum of the local correlations at every image point (Hale, 2007b). We reformulate the problem of evaluating the orthogonality between dip and displacement using local correlations (as defined in equation 3) penalized along the dip direction (Figure 3). If the velocity model is correct, the maximum of the correlation lies along the reflector slope (Figure 3(a)); otherwise, we observe a deviation that represents the relative shift of one image with respect to the other (Figure 3(b)). To measure the shift between the two images, we consider the penalty operator

\[
  P(x, \lambda) = \lambda \cdot \nu(x),
\]

where \( \lambda \) is the correlation lag vector and the dip vector \( \nu(x) \) is the normal to the reflector at \( x \). The penalty operator \( P(x, \lambda) \) is a linear function in the dip direction (normal to the reflector) and the isopenalty lines are parallel to the reflector. The penalty operator is identically zero in the direction tangential to the reflector. The local correlation of the two images is an even function with respect to the direction of the reflector: it is so because the two images illuminate the same area of the model and are therefore similar. The penalty operator changes the symmetry of the local correlation: if the velocity model is correct, the product between \( P(x, \lambda) \) and \( c(x, \lambda) \) is an odd function in the dip direction and by stacking over the correlation lags we obtain zero (Figure 3(a)); on the other hand, if the velocity model is incorrect, the maximum of the local correlation deviates from the direction of the reflector (Figure 3(b)) and is enhanced by the penalty operator. The stack over the correlation lags does not sum to zero because the function is not odd in the dip direction and we obtain a measure of relative shift between the two images. Since the maximum of the local correlation is positive, the sign of the penalty operator in the dip direction indicates the direction of the relative shift of one image with respect to the other. Figure 4 shows the local correlations, penalty operators, and penalized local correlations for a particular location on the reflector in the three cases in Figure 2. The rows of Figure 4 refer to different models (from top to bottom: a too high, the correct, and a too low velocity model). The point on the image has lateral position \( x = 3.5 \) km and vertical that depends on the imaged reflector.
Wavefield tomography based on local image correlations

Figure 1. Test model: single horizontal density interface in a constant 0.5 s/km slowness medium. The lighter shades of grey correspond to higher values of density.

Figure 4(a), 4(d), and 4(g) show the local correlations computed from two nearby shots; observe the shift in the peak of the local correlation and the orientation of the correlation panels. Figures 4(b), 4(e), and 4(h) show the penalty operators that we use to highlight the inconsistency between the dip measured on the image and the local correlations. The orientation of the penalty operators does indeed depend on the measured dip. The applied tapering reduces truncation artifacts from the edges of the local correlation panels. Figures 4(c), 4(f), and 4(i) show the penalized correlations. The apparent shift is now clear as a sign unbalance in the panels. When the velocity model is correct, the penalized correlation is an odd function (i.e., zero mean value) (Figure 4(f)); on the other hand, the panels lose their symmetry when the model is incorrect (Figures 4(c) and 4(i)). The sign of the mean value of the penalized panels measures the apparent shift between the two images.

We define the objective function

\[ J(m) = \frac{1}{2} \sum_{i} \left\| r_i(x) \right\|_x^2 \]

\[ = \frac{1}{2} \sum_{i} \left\| \int P(x, \lambda) c_i(x, \lambda) d\lambda \right\|_x^2, \]

where \( m \) represents the model parameters. Figure 5 shows the values of the objective function for different constant perturbations of the model for the simple example shown in Figure 1. The objective function is smooth and convex in the range of errors considered.

2.5 Computation of the gradient with the adjoint-state method

Because of the dimensionality of the model, the optimization problem is addressed using a local gradient-based method (steepest descent, conjugate gradient, etc.) (Vogel, 2002). The computation of the derivative of the state variables with respect to the model parameters (Fréchet derivatives) is not practical (or even possible) because of the large number of dimensions of the model space. The adjoint-state method (Lions, 1972; Plessix, 2006) is an efficient algorithm that computes the gradient of an objective function, which depends on some variables that describe the state of the physical system under analysis (state variables), without computing the Fréchet derivatives of these variables with respect to the model parameters.

The adjoint-state method consists of four steps: the computation of the state variables, the computation of the adjoint sources, the computation of the adjoint-state variables, and finally the computation of the gradient of the objective function. The first and third step (the evaluation of the state and adjoint-state variables) require the solution of the wave-equation that governs the physics of the problem. The adjoint-state method is computationally efficient because only two wavefield simulations are required at each iteration for computing the gradient of the objective function instead of a simulation for each parameter in the model. The Fréchet derivatives of the state variables with respect to the model parameters are extremely big objects: they are defined in the space given by the cartesian product between the space of the state variables and the space of the model. They are not practical to compute and impossible to store in memory for realistic case scenarios in exploration geophysics. Nonetheless, they describe the sensitivity of each state variable to changes in each model parameter and are a powerful tool for resolution analysis.

In the following, we refer to the generic state variable using the letter \( u \), the adjoint sources are indicated as \( g \), and the adjoint-state variables are designated by the letter \( a \). For wavefield tomography, there are two adjoint sources per experiment \( i \): the source wavefield \( u_{s,i}(x, t) \) and receiver wavefield \( u_{r,i}(x, t) \), where \( i \) represents the shot index.

The state variables \( u \) are obtained from the solution to the forward problem

\[ L(m) u = f, \]

where \( f \) is the the source term vector, \( L(m) \) is the forward
Figure 2. Relationship between the dip field and the apparent displacement: dip (white) and displacement (black) vector fields for different values of slowness error for a single horizontal reflector: (a) $\Delta \frac{m}{m} = 10\%$, (b) $\Delta \frac{m}{m} = 0$, and (c) $\Delta \frac{m}{m} = 10\%$. The correct slowness value is $0.5 \text{ s/km}$.
Wavefield tomography based on local image correlations

...modeling operator, and \( m \) indicates the model parameters. \( \mathcal{L} \) can be a one-way or two-way wave operator; thus we can use either downward continuation (Stolt, 1978; Gazdag and Sguazzero, 1984; Stoffa et al., 1990) or reverse-time migration (Baysal et al., 1983; McMechan, 1983) in the wavefield tomography procedure. Here, we consider a two-way wave operator (d’Alambert or Helmholtz), and \( m \) represents the squared slowness. As shown in the following, the choice of squared slowness is convenient because it simplifies the final calculation of the gradient and allows us to obtain an expression that is independent on the model parameters at the current iteration.

The adjoint sources \( g \) are the partial derivatives of the objective function with respect to the state variables:

\[
g = \frac{\partial \mathcal{J}}{\partial a}.
\]

The actual expression of the adjoint sources depends on the objective function. The particular design of the objective function impacts the adjoint sources and characterizes the particular wavefield tomography strategy proposed. In the following sections, we derive the expressions for the adjoint sources for an objective function based on local image correlations (equation 6).

We solve the adjoint problem and compute the adjoint variables \( a \) using the adjoint sources \( g \) as the force term:

\[
\mathcal{L}^* (m) a = g,
\]

where \( \mathcal{L}^* (m) \) is the operator adjoint to \( \mathcal{L} (m) \). Finally, the gradient is given by the inner product

\[
\frac{\partial \mathcal{J}}{\partial m} = -\langle a, \frac{\partial \mathcal{L}}{\partial m} \rangle.
\]

For example, for the Helmholtz operator parametrized in terms of slowness squared, \( \mathcal{L} = -\nabla^2 - mu^2 \), the partial derivative \( \frac{\partial \mathcal{L}}{\partial m} \) returns a simple scaling factor \( -\omega^2 \), and the inner product \( \langle a, \frac{\partial \mathcal{L}}{\partial m} \rangle = -\omega^2 \langle a, u \rangle \) can also be seen as the zero-lag correlation of \( u \) and \( a \), similar to the procedure employed in conventional FWI.

2.6 Adjoint sources

As described in the previous section, the adjoint sources \( g = \frac{\partial \mathcal{J}}{\partial a} \) are the derivatives of the objective function with respect to the state variables. In our objective function, we consider the dip field a slowly varying function of state variables and neglect the derivative of the penalty operator with respect to \( u \).

The details of the derivation of the expressions for the adjoint sources are in appendix A; here, we report the final result:

\[
\begin{align*}
g_{s,i} &= u_{s,i} (\eta, t) \left[ r_{i-1} \int P (x, \lambda) w \left( x - \eta + \frac{\lambda}{2} \right) R_{i-1} (\eta - \lambda) \, d\lambda \right. \\
& \quad + \left. r_i \int P (x, \lambda) w \left( x - \eta - \frac{\lambda}{2} \right) R_{i+1} (\eta + \lambda) \, d\lambda \right],
\end{align*}
\]

and

\[
\begin{align*}
g_{r,i} &= u_{r,i} (\eta, t) \left[ r_{i-1} \int P (x, \lambda) w \left( x - \eta + \frac{\lambda}{2} \right) R_{i-1} (\eta - \lambda) \, d\lambda \right. \\
& \quad + \left. r_i \int P (x, \lambda) w \left( x - \eta - \frac{\lambda}{2} \right) R_{i+1} (\eta + \lambda) \, d\lambda \right],
\end{align*}
\]

where \( g_{s,i} = g_{s,i} (x, \eta, t) \) and \( g_{r,i} = g_{r,i} (x, \eta, t) \) indicate the adjoint sources for the source and receiver wavefield of the \( i \)th shot, respectively, and \( r_i = r_i (x) \) is defined as in equation 2.

The expressions of the adjoint sources in equation 10 and 11 are quite complex and need to be explained in some detail. Notice that the adjoint sources depend on 3 variables \((x, \eta, t)\). The vector \( x \) represents the physical space, where model and image are defined; the variable \( t \) identifies the time axis; \( \eta \) is an auxiliary vector defined in the physical space that spans the local window around every image point. The adjoint source is obtained by spreading the value of the residual at each point in a window around the image point and by weighting it by the...
Figure 4. Measure of the relative displacement between two images using penalized local correlations. We consider 2 images from nearby experiments of the reflector in Figure 2; each row (from top to bottom) represents a different model: too high, correct, and too low; the columns show (from left to right) the local correlation panel, the penalty operator used, and the penalized local correlation. For each model, we pick a point on the reflector with lateral coordinate $x = 3.5$ km. The vertical coordinate changes as a function of model parameters because the depth of the reflector changes. Observe the asymmetry of the penalized local correlations for the wrong models: the mean value (i.e., the stack over the correlation lags) of the penalized correlations gives us a measure of the relative shift between two images.

Figure 5. Values of the objective function for different errors in the slowness model. We consider constant perturbations ranging from $-10\%$ to $10\%$ of the exact value ($0.5$ s/km).
the integral over $\lambda$, which represents a local convolution of the image and the penalty operator. The residual in equation 2 measures the relative displacement of one image with respect to the other; the adjoint source thus estimates the curvature of the moveout in the shot-domain common-image gathers at each image point and scales the background wavefields by this value.

Equations 10 and 11 depend on two images, $R_{i-1}(x)$ and $R_{i+1}(x)$, and may suffer from cycle skipping if the distance between the two shots and the model complexity cause the reconstituted reflectors to be more distant than a quarter of a wave-length. In order to overcome this problem, we use the central image $R_i(x)$ to compute the adjoint sources.

The dependence of the penalty operator $P(x, \lambda)$ on the state variables $u$ passes through the definition of the dip vector $\nu(x)$, which is normal to the reflectors at every point in the image. The dip field defines the wavefront set of the function that represents the reflection events in the image (Chang et al., 1987); the link between the image (and the wavefields) and the dip field can be written exploiting the concept of gradient square tensor (van Vliet and Verbeek, 1995): the dip vector is the eigenvector associated with the largest eigenvalues. There is no simple linear relationship between the dip vector and the wavefields, i.e. the state variables. Neglecting this term in the computation of the adjoint sources may introduce an error of the same order of magnitude of the term considered. A thorough study of $\partial P/\partial u$ and its impact on the computation of both the adjoint sources and the gradient of the objective function is subject for future research.

In the following test, we consider the simple density interface in Figure 1 and three shots 60 m apart starting from $x = 4$ km. The data are computed in a constant 2.0 km/s velocity medium using a two-way finite-difference scheme. We migrate the data with three different slowness models and perform sensitivity analysis for our methodology. We show the residuals, the objective function as a function of the model parameters, and the gradient with respect to the model parameters.

2.7 Flat reflector in a constant velocity medium

Figure 6 shows the migrated images of a shot located at $x = 4$ km using three different velocity models. Observe that with a single shot, the image of a horizontal reflector curves up or down depending on the sign of the velocity error. The relative displacement between different images is measured by the residual defined in equation 2. Figure 7 shows the residuals calculated using two images from adjacent shots. Observe the change in sign across the zero-offset reflection point and how the sign changes with the velocity error. The shift between the images tells us that the velocity model contains errors but it does not unequivocally relate with the sign of the error. We can combine the residuals and obtain an estimate of the curvature of the shot-domain common-image gathers (Figure 8) that is more directly connected to the sign of the velocity error. The sign flip at the edges of the reflector are due to the limited aperture of the experiment and the fact that the three images do not overlap over the entire extent of the reflector.

In Figure 9, we show the gradient computed from the correlation objective function. We consider three shots to compute the residuals and use only the wavefields of the central experiment to compute the gradient. Observe the correlation between the sign and pattern of the curvature estimate in Figure 8 and the computed gradient of the objective function. The gradient of the objective function represents the direction of maximum increase of the function itself and its sign depends on the chosen parametrization. We compute the gradient with respect to slowness squared; this means that a positive gradient points toward slower models whereas a negative gradient points toward faster models.

3 EXAMPLES

We compute the gradients for the objective function based on local image correlations using the adjoint-state method and show a few inversion tests. We discuss 3 cases of increasing complexity. First, we consider a horizontal density interface (Figure 1) in a homogeneous slowness model. We use 30 sources evenly spaced 200 m apart at the surface and receivers located at every grid point at $z = 0$. We demonstrate how the model is updated through inversion by means of a gradient-based method and the adjoint-state method. Second, we run an inversion test on a simple heterogeneous model with different layers using 40 shots evenly spaced 200 m apart and receivers at every grid point on the surface. We use shot-domain common-image gathers to assess the quality of the result and to show that our methodology is able to obtain a set of images that is invariant with respect to the experiment position. The final test uses the Marmousi model. The data are generated with a single scattering code and migrated with a downward continuation scheme.

3.1 Homogeneous slowness model

We consider again the simple model in Figure 1 with a single horizontal density interface. The data are generated using a time-domain finite difference code with absorbing boundary conditions implemented on all sides of the model. Random Gaussian noise is added to the data and the signal-to-noise ratio is equal to 10. The correct model (Figure 10) is a homogeneous slowness layer and the initial model is homogeneous but with an incorrect value of slowness (Figure 11(a)). We model 30 shots at the surface starting from $x = 1.01$ km; the spacing between the shots is 200 m and the 400 receivers are located at every grid point on the surface.

In the image obtained with the initial model (Figure 11(b)), the reflector is mispositioned because of the erroneous value of the model parameters; the shot-domain common-image gathers computed for illustration purposes show that different shots image the reflector at different depths (Figure 11(c)). We run five iterations of steepest descent (Vogel, 2002) and reconstruct the slowness model above the re-
Figure 6. We consider the simple model in Figure 1. We migrate the data with a slowness model that is too low (a), correct (b), and too high (c). The slowness squared error is constant across the entire model and equal to 10% of the background. The correct slowness value is 0.5 s/km.
Figure 7. For the three cases in Figure 6, we compute the residuals according to Equation 2. The sign of the residual indicates the relative displacement of one image with respect to the other.
Figure 8. The difference between the residuals from adjacent experiments give us an estimate of the curvature of the shot-domain common-image gathers. This information correlates well with the velocity error in the model.
Figure 9. Gradient of the correlation objective function for a slowness model that is too low (a), correct (b), and too high (c). The slowness squared error is constant everywhere and equal to 10% of the background. The distance between the two shots is 60 m. We consider three shots and computed the gradient using the wavefields of the central experiment.
flector (Figure 12(a)). The migrated image obtained with the final model (Figure 12(b)) is shifted upward as a result of the higher slowness value. The common-image gathers in Figure 12(c) show the focusing obtained with the reconstructed model. The far left and far right edges of the reflector show residual curvature, which is due to the lack of illumination and of constraints from the data. The lack of illumination leads to a higher value of slowness in the central part of the model: the inversion compensates the lower slowness on the sides with a higher slowness in the middle. Observe the rapid decrease of the objective function as the gathers are flattened (Figure 13).}

3.2 Synthetic laterally heterogeneous model

We use the synthetic heterogeneous model in Figure 14 to test our inversion algorithm. The model has a few layers with different dips and a sincline structure with an inversion in the otherwise decreasing slowness trend. We generate full-acoustic (i.e., non-Born) data with absorbing boundary conditions (no free-surface multiples) and random Gaussian noise with a signal-to-noise ratio equal to $10$. We model 40 shots evenly spaced 200 m apart and receivers at every grid point on the surface. We heavily smooth the initial model in Figure 14 and obtain the initial slowness model for migration (Figure 15(a)). Figure 15(b) shows the stack of the migrated images obtained from the initial model; if we compare the correct model in Figure 14 and the migrated image we can observe the mispositioning of the reflecting interfaces. We construct 10 shot-domain common image gathers to assess the overall focusing of the image. These gathers are simply the juxtaposition of the migrated images at fixed lateral positions (from $x = 0$ km to $x = 9$ km every 1 km). Figure 15(c) shows such gathers for the initial model: the position of the reflectors changes as a function of experiment (the images do not satisfy the semblance principle) and the model is thus not accurate. After 15 waveform tomography steps we recover the model in Figure 16(a). The model is still quite smooth but the migrated image shows a noticeable improvement in the position of the reflectors (Figure 16(b)), and the shot-domain common-image gathers in Figure 16(c) clearly indicate better focusing. Moreover, the sign of the reflectors is now consistent with the actual slowness contrasts in the exact model. We impose conformity between the model update and the layers in the image by means of a structure-oriented smoothing operator, which steers the gradient of the objective function by smoothing along the reflector slope. This approach is conceptually similar to the sparse inversion proposed by Ma et al. (2010). Figure 17 shows the evolution of the objective function with iterations. Notice the monotonic decrease in the value of the residual and the flattening when the algorithm converges. The inversion also shows a rapid convergence in the very first iterations of steepest-descent: the structure-oriented smoothing forces the gradient to conform to the geometry of the layers and speeds up convergence to an acceptable model. Despite the simplicity of the model, the inaccuracy of the wave kinematics and the absence of overturning energy (diving waves) make the data-domain techniques (such as full waveform inversion) ineffective in this particular case.

3.3 Marmousi model

We set up a more complex inversion test using the Marmousi model. We model the data using a finite-difference single-scattering (i.e., Born) modeling code and absorbing boundary conditions on the 4 sides of the model. We simulate 78 shots, the spacing between the shots is 0.08 km, and the first source is at $x = 0.96$ km; receivers are placed at every grid point at $z = 0$ km. Conflicting dips are not yet handled correctly by our methodology because they make the definition of the penalty operators ambiguous. For this reason, we restrict the inversion to the first 1.5 km in depth, where the reflectors are more coherent.

The initial model is a heavily smoothed version of the correct Marmousi slowness model. Figure 18 shows the ini-
Figure 11. Initial slowness model (a), associated migrated image (b), and shot-domain common-image gathers (c). Observe the curvature of the gathers: different experiments image the reflector at different depths.
Figure 12. Reconstructed slowness model (a), final migrated image (b), and shot-domain common-image gathers (c). The inversion involves 5 iterations of steepest descent. Observe the horizontal gathers that indicate the invariance of the image with respect to the shot position. The area of the model that is well illuminated is smaller than the actual extent of the perturbation and the inverted model has a higher value of slowness than the correct medium.
Wavefield tomography based on local image correlations

Figure 13. The objective function decreases as a function of iteration and flattens out when the common-image gathers are horizontal.

Figure 14. Slowness model used to generate the full-acoustic data. Absorbing boundary conditions have been applied at all sides of the model. The lighter shades of grey indicate faster layers. The top layer has slowness of water (0.66 s/km).

tial slowness model, the associated migrated image, and shot-domain common-image gathers extracted every 1 km from x = 0 km. Because of the regularity of the geologic structures in the shallow part of the model, the migrated image is not severely distorted, nonetheless the reflectors are mis-positioned and the faults are not correctly focused. The shot-domain common-image gathers (Figure 18(c)) show variation of the depth of imaged reflectors as function of experiment, which indicates model inaccuracy.

We run 14 iterations of steepest-descent and obtain the results in Figure 19. The slowness model is corrected by the inversion procedure and the imaged interfaces are moved toward the correct position. The fault planes are better focused and we reduce the interface crossings at the unconformities (x = 3 km, x = 4 km and x = 5 km). The shot-domain common-image gather in Figure ?? show flatter horizontal events that indicate a more accurate kinematic model, and additional events are focused at the bottom of the model (at about z = 3 km). Observe also that the common-image gathers in poorly or non illuminated parts of the model contribute mainly migration noise and thus cannot be used to assess the correctness of the velocity model.

In Figure 20, the objective function shows a constant decrease. The residuals are computed over the entire image but we use only the points in the first 1.5 km in depth to update the velocity model. For complex areas with conflicting dips and complicated geologic features (where there is no clear definition of the dip field), a more sophisticated design of the penalty operator is key to obtaining meaningful residuals and a reliable gradient. The eigenvalues and eigenvectors of the gradient square tensors (van Vliet and Verbeek, 1995) can be used
Figure 15. Initial slowness model (a), associated migrated image (b), and shot-domain common-image gathers. The reflectors are severely defocused and mispositioned. Observe the curvature of the common-image gathers that shows how different experiments image the reflectors at different depths.
Figure 16. Slowness model after 15 iterations of waveform inversion (a), migrated image (b), and shot-domain common-image gathers (c). Observe the improved focusing and positioning of the reflectors. The common-image gathers show that the images are invariant in the shot direction and the model is thus kinematically correct. Observe the profile of the various layers in the reconstructed model.
to define ellipses that, in turn, may offer a structure oriented criterion for the definition of the penalty operators.

4 DISCUSSION

A complete image of the subsurface is the superposition of partial images from individual experiments. The semblance principle (Al-Yahya, 1989) is one common criterion for assessing the correctness of the velocity model used for imaging the survey: when the velocity model is correct, all shots locate reflectors at the same position, i.e., the image is invariant along the experiment axis. Several shots are needed to evaluate a velocity error at a single point in space.

We can evaluate the invariance along the experiment axis by computing the energy of the first derivative in that dimension. The first derivative acts as a penalty operator by highlighting and enhancing deviations from the horizontal direction along the shot axis. The stack of the energy of the first derivative is the differential semblance operator applied directly in the shot domain (Symes, 1991; Plessix, 2006).

Here, we explore an alternative statement of the semblance principle: when the velocity model is correct, images from different shots constructively interfere and build up the image perpendicular to the structural dip or, equivalently, parallel to the reflector slope, at every point in the image space. The structural dip is a commonly extracted attribute and can be linked to the image itself by means of the gradient square tensors (van Vliet and Verbeek, 1995). Unfortunately, the relationship between wavefields and the dip field is nonlinear: the dip vector represents the eigenvector associated with the largest eigenvalue of the gradient square tensor. The integration of information about the dip variation with respect to model perturbation is not straightforward, and further research is needed to develop an efficient method to exploit this information for velocity analysis purposes. Nonetheless, we are able to measure the semblance of two images through appropriately penalized local correlations of pairs of images. If the velocity model is correct, the maximum of the local correlation is along the reflector at every point in the image; if the model is incorrect, the maximum deviates from the reflector slope. The penalty operator is space-dependent and annihilates the correlation panels orthogonal to the reflector dip. Because of the dependency on the velocity model, we measure the dip field at each tomographic iteration; the estimation can be carried out efficiently using gradient square tensors (van Vliet and Verbeek, 1995; Hale, 2007a).

The correlation objective function is superior to the image difference in many respects. First, the difference between two images depends on image amplitudes that change as a function of the shot position and cannot be matched point-wise as in the standard implementation of full-waveform inversion in the data space. The amplitude patterns affect the residual in the image space and effectively contribute to the adjoint source calculation, even if the velocity model is correct and the gradient of the objective function is zero (Mulder and Kroode, 2002). By penalizing local correlations, we reduce the dependence of the objective function on amplitudes, thus increasing the robustness and reducing the systematic bias caused by the amplitude differences between the images. A downside of the correlation operator is the loss of spatial resolution. Correlation in the spatial domain is equivalent to multiplication in the dual frequency/wavenumber domain; for signals with both finite spatial support and bandwidth, the multiplication of the spectra decreases the bandwidth, i.e., increases the width of the signal in the spatial domain. To increase the resolution and accuracy of the evaluation of the relative shifts between images, deconvolution is a viable improvement over correlation. If we assume that two images from nearby experiments are
Wavefield tomography based on local image correlations

Figure 18. Initial slowness model (a), initial migrated images (b), and shot-domain common-image gathers (c). The initial model is obtained by heavily smoothing and scaling by 0.9 the correct model.

linked by a simple spatial shift, a local deconvolution approach would ideally produce a bandlimited spatial delta function indicating the direction of apparent displacement. On the other hand, since deconvolution amplifies the noise in the data (because of the division in the frequency domain), additional care is necessary to stabilize the result.

A thorough study of the effect of the shot distance on the computed gradient would be beneficial in relation to phenomena such as cycle skipping, which hinders many velocity analysis strategies (especially in the data domain but not exclusively). The image domain is intrinsically less prone to cycle skipping problems; nevertheless the question remains about what happens when two images illuminate rather different areas of the subsurface and the local correlations cannot be used to estimate a reliable and meaningful shift between the two images. From our current knowledge and understanding, the shots must be close enough to provide comparable images of the subsurface and avoid cycle skipping in the image domain.

The sampling of the shot position and the illumination pattern in the subsurface can also create a scenario in which the two images do not overlap in certain local windows.

Our method is based on locally coherent events, such as locally smooth reflectors. Conflicting dips, fault planes, and areas where the definition of a reflection plane is ambiguous represent open problems because in these areas we cannot define penalty operators. In this respect, plane-wave migration may represent a valid solution because of the implicit spatial filtering of the image: each plane-wave reconstructs a particular subset of the dips in the model, thus reducing the ambiguity in defining the penalty operators.

In contrast with other techniques in the image space (Bishop et al., 1985; Biondi and Symes, 2004; Lambaré et al., 2004; Xie and Yang, 2008; Yang and Sava, 2010, 2011), we do not need to compute image gathers (or extended images). A further cost-saving factor is given by the selection of points on the imaged reflectors (Yang and Sava, 2010) and away from
complex areas (pinch-outs or areas with conflicting dips). The computation of the local correlation is carried out at every image point using the efficient method developed by Hale (2006).

The ability to extract velocity information from pairs of experiments adds a new degree of freedom for implementing a model update procedure. Here, we indicate two possible strategies. We can simultaneously include all the shots in the survey in the definition of the objective function or we can proceed iteratively and update the velocity model using the information obtained from a single pair of experiments before moving to the next pair. Nearby shots probe similar portions of the model and provide comparable images; we can use the information extracted from an initial group to update the model used for imaging a second group of experiments. Since in any migration velocity analysis scheme we have to image the entire survey at least once, the iterative update of the model over shots becomes cost-effective if we actually reduce the number of global migrations of the entire dataset.

In 2D, we can easily define an order for the experiments; in the general 3D scenario, we have one extra degree of freedom. The definition of the objective function remains the same but a given shot can have more than 2 neighboring experiments (the acquisition is defined on a bidimensional grid). We can generalize the concept of correlation using the semblance functional (Taner and Koehler, 1969), which is nonlinear in the input signals and makes the computation of the gradient more involved. We can also separately analyze each pair of experiments given a reference model. Further research and numerical tests are needed to assess efficiency and robustness of one strategy over the other.

5 CONCLUSION

We develop an objective function for migration velocity analysis in the shot-image domain. Our methodology is based on local image correlations and a reformulation of the semblance principle involving only small groups of migrated images. The optimization criterion is the minimization of the apparent shift
between the migrated images and naturally leads to a differential semblance optimization problem, which is iteratively solved using a gradient-based method. The linearity of the operators in the objective function makes the computation of the gradient practical using the adjoint-state method. Our approach is full-wave because we do not rely on the linearization of the wave extrapolation procedure with respect to the model parameters to construct a migration velocity analysis operator. We iteratively solve a nonlinear problem with a gradient-based algorithm and the simulation of the wavefields is fully nonlinear with respect to the model parameters. We test our method using a few synthetic examples of increasing complexity.

The structural features in the migrated image and the variation of these features as function of experiment supply valuable information about the model parameters. Here, we show that it is possible, without constructing common-image gatherers, to extract information about the model from a very limited number of experiments using a warping relationship between migrated images. Despite the fact that we are neglecting the variation of the image dip field with respect to the model parameters, we are nevertheless able to correct anomalies in the model.

6 ACKNOWLEDGMENTS

We would like to thank eni E&P for the permission to publish this work and Nicola Bienati for the valuable conversations on tomography and inversion. The reproducible numerical examples in this paper use the Madagascar open-source software package freely available from http://www.reproducibility.org. This research was supported in part by the Golden Energy Computing Organization at the Colorado School of Mines using resources acquired with financial assistance from the National Science Foundation and the National Renewable Energy Laboratory.

7 A. COMPUTATION OF THE SOURCE AND RECEIVER ADJOINT SOURCE

Our objective function is defined as

\[
\mathcal{J}(m) = \frac{1}{2} \sum_i \| r_i(x) \|^2_x
\]

\[
= \frac{1}{2} \sum_i \left\| \int P(x, \lambda) c_i(x, \lambda) d\lambda \right\|^2_x, \tag{A-1}
\]

where \( c_i(x, \lambda) = \int_{w(x)} r_{i+1}(\xi - \frac{1}{2}) R_i(\xi + \frac{1}{2}) d\xi \) is the local correlation of two images from the nearby shots \( i \) and \( i + 1 \), and \( P(x, \lambda) \) is an operator that annihilates the local correlation panel along the direction of the structure at every point in the image.

The adjoint-state method computes the gradient of the objective function with respect to the model parameters by solving two forward problems for each state variable of the system. This procedure turns out to be extremely computationally efficient since it avoids perturbing each model parameter and computing the resulting state variable perturbation (Lions, 1972; Plessix, 2006). The solution of the auxiliary forward problem requires different sources (adjoint sources) that are computed by differentiating the objective function with respect to the state variables.

We now derive the mathematical expression for the source and receiver adjoint sources. The computation of the derivative of a functional with respect to a function requires the definition of an inner product in the appropriate functional space; moreover, it requires the concept of the Fréchet differential (Vogel, 2002).

Let us denote the state variables of the problem, i.e., the source and receiver wavefields for the \( i \)th shot, by \( u_{s,i} \) and \( u_{r,i} \), respectively. We derive the expression of the adjoint source for the source wavefield. The derivation and expression...
of the adjoint source for the receiver wavefield are analogous. First, we make explicit the dependency of the residual \( r_i (x) \) (and thus of the objective function \( \mathcal{J} \)) on the state variable: 

\[
\delta r_i (x) = \int P(x, \chi) w(x - \xi) u_{s,i}(\xi - \frac{\lambda}{2}, t) \cdot u_{r,i}(\xi - \frac{\lambda}{2}, t) R_{i+1}(\xi + \frac{\lambda}{2}) dV, \tag{A-2}
\]

where the integral is a triple integral over the volume element \( dV = dt d\xi d\lambda \). Second, we compute the Fréchet differential of the objective function \( \mathcal{J}(\eta) \) with respect to an arbitrary perturbation of the state variable \( \delta u_{s,i} \). The objective function in equation A-1 is the sum over shots of energy of the residues in equation A-2 over the spatial domain. In order to simplify the derivation, we first write the differential of the objective function as a function of the differential of the residual using perturbation theory:

\[
\mathcal{J}[u_{s,i}] + \delta \mathcal{J}[u_{s,i}, \delta u_{s,i}] = \frac{1}{2} \int (r_{i+1} - \delta r_{i+1}[u_{s,i}, \delta u_{s,i}])^2 dV + \int r_{i+1} \delta r_i[u_{s,i}, \delta u_{s,i}] dV \tag{A-3}
\]

Expanding the squares in equation A-3 and considering only the terms that are linear in the wavefield perturbation, we obtain an expression for the Fréchet differential of the objective function:

\[
\delta \mathcal{J}[u_{s,i}, \delta u_{s,i}] = \int r_{i+1} \delta r_{i+1}[u_{s,i}, \delta u_{s,i}] + r_i \delta r_i[u_{s,i}, \delta u_{s,i}] dV. \tag{A-4}
\]

The square brackets in equation A-4 indicate the double dependency of the Fréchet differential on the state variable and its arbitrary perturbation.

Assuming \( r_i (x) \) is Fréchet differentiable, we compute the Fréchet differential using the limit of the incremental ratio:

\[
\delta r_i[u_{s,i}, \delta u_{s,i}] = \lim_{h \to 0} \frac{r_i[u_{s,i} + h\delta u_{s,i}] - r_i[u_{s,i}]}{h}, \tag{A-5}
\]

where \( h \) is a scalar, and \( \delta u_{s,i} \) is an arbitrary perturbation of the state variable \( u_{s,i} \) and belongs to the same functional space. The residual \( r_i \) is a linear functional of \( u_{s,i} \), and computing the limit in equation A-5 we obtain

\[
\delta r_i = \int P(x, \chi) w(x - \xi) \delta u_{s,i}(\xi - \frac{\lambda}{2}, t) \cdot u_{r,i}(\xi - \frac{\lambda}{2}, t) R_{i+1}(\xi + \frac{\lambda}{2}) dV. \tag{A-6}
\]

We want to express equation A-6 as an inner product between an operator (the Fréchet derivative) and the state variable perturbation \( \delta u_{s,i} \); we perform a change of variables to make the perturbation of the state variable \( \delta u_{s,i} \) depend only on a single independent variable. We define two new variables \( \eta = \xi - \frac{\lambda}{2} \) and \( \chi = \frac{\lambda}{2} \); the Jacobian of this transformation is trivial and equal to 1; then, in the new coordinate system, we have

\[
\delta r_i = \int P(x, \chi) w(x - \eta - \frac{\lambda}{2}) \delta u_{s,i}(\eta, t) \cdot u_{r,i}(\eta + \frac{\lambda}{2}) R_{i+1}(\eta + \frac{\lambda}{2}) dV', \tag{A-7}
\]

where the triple integral is over the new volume element \( dV' = dt d\eta d\chi \). From equation A-7, we find the expression of the gradient of the residual \( r_i \) with respect to the state variable \( u_{s,i} \):

\[
\nabla_{u_{s,i}} r_i = \int P(x, \chi) w(x - \eta - \frac{\lambda}{2}) u_{r,i}(\eta, t) \cdot R_{i+1}(\eta + \frac{\lambda}{2}) d\chi. \tag{A-8}
\]

Combining equation A-8 into A-4, we can finally write the equation for the gradient of the objective function \( \mathcal{J} \) with respect to the state variable \( u_{s,i} \) as

\[
\nabla u_{s,i} \mathcal{J} = \int r_{i+1} P(x, \chi) * R_{i+1} (x) \tag{A-9}
\]

where

\[
P(x, \chi) * R_{i+1} (x) = \int w(x - \eta + \frac{\lambda}{2}) \cdot P(x, \chi) R_{i+1} (\eta - \chi) d\chi
\]

and

\[
P(x, \chi) * R_{i+1} (x) = \int w(x - \eta - \frac{\lambda}{2}) \cdot P(x, \chi) R_{i+1} (\eta + \chi) d\chi
\]

are the local convolution and local correlation of the penalty operator with the image over the local window \( w \), respectively. The derivation for the receiver-side adjoint source is totally analogous and leads to the formula

\[
\nabla u_{r,i} \mathcal{J} = \int u_{s,i}(\eta, t) [r_{i+1} P(x, \chi) * R_{i+1} (x) \]

\[
\tag{A-10}
\]

Equations A-9 and A-10 are the expressions for the adjoint sources in equations 10 and 11 in the paper.

REFERENCES


Wavefield tomography based on local image correlations

next leap forward in imaging at valhall: First Break, 28, 65–70.
——–, 2005, Inverse problem theory and methods for model parameter estimation: SIAM.
Vogel, C., 2002, Computational methods for inverse problems: SIAM.
Image-domain wavefield tomography with extended common-image-point gathers

Tongning Yang and Paul Sava

ABSTRACT

Waveform inversion is a velocity-model-building technique based on full waveforms as the input and seismic wavefields as the information carrier. Conventional waveform inversion is implemented in the data-domain. However, similar techniques referred to as image-domain wavefield tomography can be formulated in the image domain and use seismic image as the input and seismic wavefields as the information carrier. The objective function for the image-domain approach is designed to optimize the coherency of reflections in extended common-image gathers. The function applies a penalty operator to the gathers, thus highlighting image inaccuracies arising from the velocity model error. Minimizing the objective function optimizes the model and improves the image quality. The gradient of the objective function is computed using the adjoint-state method in a way similar to that in the analogous data-domain implementation. We propose an image-domain velocity-model building method using extended common-image-point gathers constructed at discrete locations in the image. Such gathers have the advantage over conventional common-image gathers that they are robust for imaging reflectors with a wide range of dips. The common-image-point gathers can extract the velocity information from steep reflectors imaged with a two-way wave propagator, this information improving accuracy of the gradient computation and vertical resolution of velocity estimates. The gathers moreover are effective in reconstructing the velocity model in complex geologic environments and can be used as an economical replacement for conventional common-image gathers in wave-equation tomography. A test on the Marmousi model illustrates successful updating of the velocity model using common-image point gathers and resulting improved image quality.

Key words: wavefield tomography, wave-equation, extended images, adjoint state method

1 INTRODUCTION

Building an accurate and reliable velocity model remains one of the biggest challenges in current seismic imaging practice. In regions characterized by complex subsurface structure, prestack wave-equation depth migration (e.g., one-way wave-equation migration or reverse-time migration) is a powerful tool for accurately imaging the Earth’s interior (Gray et al., 2001; Etgen et al., 2009). Because these migration methods are very sensitive to model errors, their widespread use significantly drives the need for high-quality velocity models because these migration methods are very sensitive to model errors (Symes, 2008; Woodward et al., 2008; Virieux and Operto, 2009).

Waveform inversion (WI) represents a family of techniques for velocity model building using seismic wavefields (Tarantola, 1984; Woodward, 1992; Pratt, 1999; Sirgue and Pratt, 2004; Plessix, 2006; Vigh and Starr, 2008a; Plessix, 2009; Symes, 2009). This type of methodology, although usually regarded as one of the costliest for velocity estimation, has been gaining momentum in recent years, mainly because of its accuracy as well as advances in computing technology. The core of WI is using a wave equation (typically constant-density acoustic) to simulate wavefields as the information carrier. Usually WI is implemented in the data domain by adjusting the velocity model such that simulated and recorded data match (Tarantola, 1984; Pratt, 1999). This match is based on the strong assumption that the wave equation used for simulation is consistent with the physics of the Earth. This, however,
is unlikely to be the case when the Earth is characterized by strong (poro)elasticity. In data domain approaches, significant effort is often directed toward removing the components of the recorded data that are inconsistent with the assumptions used.

Velocity-model-building methods using seismic wavefields can be implemented in the image domain rather than in the data domain. Instead of minimizing the data misfit, the techniques in this category update the velocity model by optimizing the image quality, which is the cross-correlation of wavefields extrapolated from the data and from the source wavelet. The image quality is optimized when the data are migrated with the correct velocity model, as stated by the semblance principle (Al-Yahya, 1989; Yilmaz, 2001). The common idea is to optimize the coherency of reflection events in common-image gathers (CIGs) via velocity-model-updating. Since images are obtained using full waveforms, and velocity estimation also employs seismic wavefields as the information carrier, these techniques can be considered as a particular type of WI, and we refer to them as image-domain wavefield tomography (WT). Unlike traditional image-domain ray-based tomography methods, image-domain WT uses lamp-limited wavefields in the optimization procedure. Thus, this technique is capable of handling complicated wave propagation phenomena such as multi-pathing in the subsurface. In addition, the band-limited character of the wave-equation engine more accurately approximates wave propagation in the subsurface and produces more reliable velocity updates than do ray-based methods.

Sava and Biondi (2004a, b) describe the concept of wave-equation migration velocity analysis, which is one variation of image-domain WT. The method linearizes the downward continuation operator and establishes a linear relationship between the model perturbation and image perturbation. The model is inverted by exploiting this linear relationship and minimizing the image perturbation. Sava et al. (2005) demonstrate application of the technique to velocity model building in complex regions, and Sava and Vlad (2008) discuss its detailed numerical implementation. This methodology, however, is limited by its reliance on the one-way wave propagation operator, which constrains its ability to produce model updates in complicated geology with steep reflectors.

Differential semblance optimization (DSO) is another variation of image-domain WT. The essence of the method is to minimize the difference of any given reflection between neighboring offsets or angles. Symes and Carazzone (1991) propose a criterion for measuring coherency within offset gathers and establish the theoretic foundation for DSO. The concept is then generalized to space-lag (subsurface-offset) and angle-domain gathers (Shen and Calandra, 2005; Shen and Symes, 2008). In practice, space-lag gathers (Rickett and Sava, 2002; Shen and Calandra, 2005) and angle-domain gathers (Sava and Fomel, 2003; Biondi and Symes, 2004) are two popular choices among various types of CIGs used for velocity analysis. These gathers are obtained by wave-equation migration and have fewer artifacts usually found in conventional offset gathers obtained by Kirchhoff migration, and thus they are suitable for applications in complex earth models (Stolk and Symes, 2004).

With recent developments in forward modeling and computing hardware, reverse-time migration (RTM) has become a common tool for imaging applications, especially in complex subsurface areas. One can characterize the wave propagation in the subsurface for the velocity estimation process more accurately using a two-way wave-equation propagator than using an one-way wave-equation propagator (Mulder, 2008). Furthermore, the capability of RTM for imaging steep reflections benefits velocity model building since more information can be extracted from the image to constrain the velocity updates (Gao and Symes, 2009). To effectively access the velocity information contained in steep reflections, Sava and Vasconcelos (2011) and Vasconcelos et al. (2010) propose common-image-point gathers (CIPs) as an alternative to space-lag or angle-domain gathers. CIPs are sparsely distributed in the subsurface on reflections and offer several advantages in the context of velocity inversion. First, the construction of a complete lag vector (space lags and time lag) avoids the bias toward nearly horizontal reflections. Thus, the gathers are sensitive to the velocity information in reflections with arbitrary dip and take advantage of the steep events imaged by RTM. Second, the discrete sampling of the gathers provides a flexible way to extract the velocity information from the image and facilitates target-oriented velocity updates. Furthermore, the sparse construction of the gathers reduces computational cost and storage requirements, both important in 3D applications.

In this paper, we propose an image-domain wavefield-based velocity-model-building approach using CIPs as the input. One key component of image-domain WT is wavefield simulation using a one-way or two-way wave-equation engine, similar to data-domain WI but with more flexibility. Another key component of the method is the objective function (OF), which is constructed by applying a penalty operator to CIPs whose minimization allows us to optimize image coherency and to update the velocity model simultaneously. The third component is an effective gradient calculation based on the adjoint state method (Plessix, 2006; Symes, 2009). In summary, gradient calculation with this method consists of the following steps: (1) compute the state variables, i.e., the seismic wavefields obtained from the source by forward modeling and from the data by backward modeling; (2) compute the adjoint source, i.e., a calculation based on the OF and on the state variables; (3) compute the adjoint state variables, i.e., the seismic wavefields obtained from the adjoint source by backward modeling; (4) compute the gradient using the state and adjoint state variables. We provide more details on this technique in the body of the paper.

This paper starts with a theoretical discussion of image-domain WT and its implementation with CIPs and CIGs. We show that CIPs overcome the bias toward nearly horizontal reflectors, which is important in the presence of steeply dipping structures because the information extracted from steep reflections provides additional constraints on the velocity model building. We use the Marmousi model to demonstrate that CIPs can be an economical and accurate replacement for CIGs.
Wavefield tomography with CIPs

used in the conventional wave-equation-based DSO approach to model building in complex subsurface areas. The results obtained from CIPs are comparable to those obtained from CIGs, but with smaller cost for computing and storing the image gathers.

2 THEORY

In this section, we formulate image-domain wavefield tomography using both space- and time-lag extended images (extended CIPs) or space-lag extended images (also known as subsurface-offset CIGs). The gradient is computed by applying the adjoint-state method (Plessix, 2006; Symes, 2009), which is also a common practice for data-domain full-waveform inversion (Tarantola, 1984; Sirgue and Pratt, 2004; Virieux and Operto, 2009). This approach can easily be generalized to other image-domain wavefield tomography methods implemented with different input image gathers, e.g., time-lag CIGs (Yang and Sava, 2011).

For simplicity, we discuss the methodology in the frequency-domain rather than in the time-domain although the latter is completely equivalent and analogous. We formulate the inverse problem by first defining the state variables, through which the OF is related to the model parameters. The state variables for our problem are the source and receiver wavefields $u_s$ and $u_r$, obtained by solving the following acoustic wave equation:

$$
\begin{bmatrix}
L(x, \omega, m) & 0 \\
0 & L^*(x, \omega, m)
\end{bmatrix}
\begin{bmatrix}
U_s(j, x, \omega) \\
U_r(j, x, \omega)
\end{bmatrix} =
\begin{bmatrix}
F_s(j, x, \omega) \\
F_r(j, x, \omega)
\end{bmatrix},
$$

(1)

where $F_s$ is the source function, $F_r$ are the recorded data, $j = 1, ..., N_s$, where $N_s$ is the number of shots, $\omega$ is the angular frequency, and $x = \{x, y, z\}$ are the space coordinates. The wave operator $L$ and its adjoint $L^*$ propagate the wavefields forward and backward in time, respectively, using either a one-way or two-way wave equation. In this formulation, we designate the operator $L$ to be

$$L = -\omega^2 m - \Delta,$$

(2)

where $\Delta$ is the Laplace operator, and model parameter $m$ represents slowness squared.

In the second step of the adjoint-state method, we first construct the OF and then the adjoint sources that are used to model the adjoint-state variables required by the gradient computation. The OF for image-domain wavefield tomography is defined using the semblance principle (Yilmaz, 2001) and measures the image incoherency caused by the model errors. Therefore, the inversion process of minimizing the OF simultaneously reconstructs the model and improves the image quality.

We consider the objective function in the $\lambda - \tau$ domain ($\lambda$ is a vector that pertains to space-lags in 2D or 3D space and $\tau$ pertains to time-lag) and use extended CIPs as the input to analyze and optimize the velocity model. Extended CIPs are obtained by applying the nonzero space- and time-lag cross-correlation imaging condition (Sava and Fomel, 2006) to the wavefields at selected points in the image. Sava and Vasconcelos (2011) analyze the kinematic characteristics of reflections in extended CIPs and point out that reflections focus at zero space- and time-lags when the migration velocity is correct. Therefore, the OF based on extended CIPs is defined as

$$\mathcal{H}_\lambda,\tau = \frac{1}{2} || K_I (x) P(\lambda, \tau) r(x, \lambda, \tau) ||_{L^2},$$

(3)

where $P(\lambda, \tau)$ is defined below, and $r(x, \lambda, \tau)$ are space- and time-lag extended images:

$$r(x, \lambda, \tau) = \sum_j \sum_\omega u_s(j, x - \lambda, \omega) u_r(j, x + \lambda, \omega) e^{2i\omega\tau},$$

(4)

the overline represents complex conjugate. The operator $T$ represents the space shift applied to the wavefields and is defined by

$$T(\lambda) u(j, x, \omega) = u(j, x + \lambda, \omega),$$

(5)

The mask operator $K_I(x)$ restricts the construction of extended images to chosen discrete locations only, such that the CIPs are constructed only on reflectors and are sampled sparsely in the subsurface. For two reasons, CIPs provide an effective and efficient way to extract velocity information from migrated images. First, no gathers are computed in areas without reflections so the computational cost can be reduced. Second, CIPs can be computed on steep reflectors where conventional CIGs fail to access the velocity information contained in these reflections. An example of a penalty operator $P(\lambda, \tau)$ for vector lags is

$$P(\lambda, \tau) = \sqrt{\lambda \cdot q^2 + (V\tau)^2},$$

(6)

Here $q$ is a unit vector in the reflection plane and $V(x)$ represents the local migration velocity. The operator $P(\lambda, \tau)$ penalizes energy away from zero space- and time-lags, which indicates the existence of velocity errors. Hence, the defocused energy outside zero space-lag is enhanced by the operator $P(\lambda, \tau)$ and forms a residual that is the basis for optimization. Figures 1(a) and 1(b) show penalty operators for CIPs on horizontal and vertical reflectors, respectively. The penalty operator defined in equation 6 represents a cylinder oriented normal to the reflector in the $\lambda - \tau$ space. If we consider only the case of horizontal space-lags, the penalty operator can be simplified as

$$P(\lambda, \tau) = \sqrt{\lambda^2 + (V\tau)^2},$$

(7)

where the space-lag vector $\lambda = \{\lambda_x, \lambda_y, 0\}$.

Given $\mathcal{H}_\lambda,\tau$ in equation 3, the adjoint sources are computed as OF’s derivatives with respect to the state variables $u_s$ and $u_r$ (Shen and Symes, 2008). In this case, the adjoint sources $g_s$ and $g_r$ are complicated because the complete lags
are involved in the computation.

\[ g_r (j, x, \omega) = \sum_{\lambda, \tau} T (\lambda) P (\lambda, \tau) K_f (x) K_r (x) P (\lambda, \tau) \]

\[ - r (x, \lambda, \tau) T (\lambda) u_r (j, x, \omega) e^{-2i\omega \tau} \]

\[ g_s (j, x, \omega) = \sum_{\lambda, \tau} T (-\lambda) P (\lambda, \tau) K_f (x) K_r (x) P (\lambda, \tau) \]

\[ - r (x, \lambda, \tau) T (-\lambda) u_s (j, x, \omega) e^{2i\omega \tau} \]  

(8)

The adjoint state variables \( a_s \) and \( a_r \) are the wavefields obtained by backward and forward modeling respectively, using the corresponding adjoint sources defined in equation 8:

\[ \begin{bmatrix} \mathcal{L}^* (x, \omega, m) & 0 \\ 0 & \mathcal{L} (x, \omega, m) \end{bmatrix} \begin{bmatrix} a_s (j, x, \omega) \\ a_r (j, x, \omega) \end{bmatrix} = \begin{bmatrix} g_s (j, x, \omega) \\ g_r (j, x, \omega) \end{bmatrix} \]  

(9)

and \( \mathcal{L} \) and \( \mathcal{L}^* \) are the same wave propagation operators used in equation 1.

The last step of the gradient computation is simply the correlation between state variables and adjoint state variables (Plessix, 2006):

\[ \frac{\partial H_{\lambda, \tau}}{\partial m} = \sum_j \sum_\omega \frac{\partial \mathcal{L}}{\partial m} \begin{bmatrix} u_s (j, x, \omega) a_s (j, x, \omega) + u_r (j, x, \omega) a_r (j, x, \omega) \end{bmatrix} \]  

(10)

where \( \frac{\partial \mathcal{L}}{\partial m} \) is the partial derivative of the wave propagation operator with respect to the model parameter. Using the definition of \( \mathcal{L} \) in equation 2, we find that \( \frac{\partial \mathcal{L}}{\partial m} \) is simply \(-\omega^2\).

In equation 10 Note that the gradient for image-domain waveform tomography consists of two correlations because we define both the source and receiver wavefields as the state variables. In contrast, the gradient computed in the data-domain approach involves only one correlation on the source side because we use only the simulated wavefield as the state variable.

An alternative to image-domain waveform tomography is using space-lag CIGs, just as for conventional DSO (Shen and Symes, 2008). We reformulate the derivation to highlight the similarity between these two approaches. The state variables are also the source and receiver wavefields \( u_s \) and \( u_r \) simulated by equation 1. The OF based on space-lag CIGs (as opposed to space- and time-lag CIPs) is

\[ H_{\lambda} = \frac{1}{2} \| K_f (x) \ P (\lambda) \ r (x, \lambda) \|^2_{\lambda, \lambda} , \]

(11)

where

\[ r (x, \lambda) = \sum_j \sum_\omega u_s (j, x - \lambda, \omega) u_r (j, x + \lambda, \omega) \]

\[ = \sum_j \sum_\omega T (-\lambda) u_s (j, x, \omega) T (\lambda) u_r (j, x, \omega) . \]

(12)

The penalty operator \( P (\lambda) \) annihilates the focused energy at zero lag and highlights the energy of residual moveout at nonzero lag (Shen and Symes, 2008):

\[ P (\lambda) = |\lambda| . \]

(13)

In practice, it is common to restrict the space-lags in the horizontal directions only, i.e., \( \lambda = \{ \lambda_x, \lambda_y, 0 \} \). The OF \( H_{\lambda} \) is minimized when the reflections focus at zero lag, an indication of the correct velocity model. This, however, occurs only when the subsurface is well illuminated. Otherwise, imbalanced illumination can result in defocusing in the gatherers even if the velocity model is correct (Yang et al., 2012). To mitigate the negative influence of the uneven subsurface illumination, we can employ a weighting function that gives low weight on the
gathers in poor illumination areas and puts more weight on the gathers in good illumination areas. As a result, the gathers that are defocused where illumination is poor contribute less to the velocity updates.

Similar to the previous case, the adjoint sources are computed as the derivatives of the OF $\mathcal{H}_a$ in equation 11 with respect to the state variables $u_s$ and $u_r$:

$$
\begin{align*}
    g_s (j, x, \omega) &= \sum_\lambda T (\lambda) P (\lambda) K_I (x) \bar{K}_I (x) \\
    &\quad r (x, \lambda) T (\lambda) u_s (j, x, \omega) \\
    g_r (j, x, \omega) &= \sum_\lambda T (-\lambda) P (\lambda) K_I (x) \bar{K}_I (x) \\
    &\quad r (x, \lambda) T (-\lambda) u_s (j, x, \omega).
\end{align*}
$$

The adjoint state variables $a_s$ and $a_r$ are computed using equation 9 given the adjoint sources defined in equation 8, and the gradient corresponding to the OF in equation 11 is computed as in equation 10.

The derivation above shows the construction of OF and detailed gradient computation for image-domain wavefield tomography. Given these two components, the solution to the inverse problem is found by minimizing the OF using nonlinear gradient-based iterative methods. In each iteration, the gradient is computed and the model update is calculated by line search in the steepest descent or conjugate gradient directions. (Vigh and Starr, 2008b) This procedure is similar to that in data-domain waveform inversion.

3 EXAMPLES

In this section, we illustrate our method with two synthetic examples and emphasize the advantages of using CIPs for velocity model building. The first example highlights the robustness of CIPs in the presence of steeply dipping reflectors. The second example demonstrates the ability of CIPs to reconstruct velocity models in complex subsurface regions.

The first synthetic model is shown in Figure 2(a), and the initial model is just the vertical gradient extended to the entire model. Figures 2(b)-2(c) show the images migrated using RTM with correct and initial velocities, respectively. The lack of 2D circular low-velocity anomaly in the initial model causes defocusing and image triplications.

To highlight the robustness of CIPs, we construct both the CIGs and CIPs in the subsurface at the positions indicated by the vertical lines and dots in Figure 2(c). Figures 3(a)-3(d) and Figures 4(a)-4(d) plot the CIGs and CIPs constructed on different reflectors for correct and initial velocities. Figures 3(c)-3(d) and 4(c)-4(d) compare the CIGs and CIPs constructed on the horizontal reflector. Here, both CIGs and CIPs correctly characterize the reflection as the gathers show either focused reflections or residual movement depending on the model used for imaging. Thus, one can assess the accuracy of the velocity models by analyzing the focusing information in the gathers. In contrast, Figures 3(a)-3(b) and 4(a)-4(b) show that for the vertical reflector, only CIPs are able to correctly characterize the reflection and thus provide velocity information for the model building. The CIGs are contaminated by artifacts because we construct the gathers using horizontal lags for vertical reflections. The reflections are sampled at every depth level in the gathers, and thus their focusing cannot be correctly characterized. The difference between CIPs and CIGs is caused by the fact that both vertical and horizontal space-lags are used in CIPs while only the horizontal space-lag is used to construct CIGs. Therefore, CIGs are biased towards reflectors with small dips and fail to evaluate the velocity information available in steep reflectors. For CIPs, one can evaluate the model accuracy by analyzing focusing of vertical and horizontal reflectors in the $\lambda_2 - \tau$ and $\lambda_2 - \tau$ panels, respectively: the vector space-lags and time-lag of CIPs remove the directional bias toward horizontal reflectors, and CIPs are robust and able to analyze the velocity information from reflections regardless of the subsurface structure.

Next, we use the adjoint-state method as described in the previous section to calculate the gradient for three different scenarios: using CIPs on the horizontal reflector only, using CIPs on the vertical reflector only, and using CIPs on both reflectors. Figure 5(a) plots the difference between the correct and initial velocities. Figures 5(b)-5(d) show the gradient of the objective function computed using different groups of CIPs, indicated by the black dots. Observe that the gradients highlight the target in different ways. The gradient in Figure 5(b) constrains the model variation in the horizontal direction. In contrast, the gradient in Figure 5(c) constrains the model variation in the vertical direction. When the gradient is computed from CIPs picked on both vertical and horizontal reflectors, as shown in Figure 5(d), the variation is more effectively controlled in both the vertical and horizontal directions.

Using the gradient computed from all CIPs, we reconstruct the velocity model shown in Figure 6(a). The image migrated with this updated model is shown in Figure 6(b). The coherency of both the vertical and horizontal reflectors is improved because of the more accurate gradient used in the inversion. This indicates that using CIPs, especially those sampled on the steeply dipping reflections, we can extract additional information from the migrated image and offer more constrains for the model building.

To show the performance of CIPs in wavefield tomography, we consider the synthetic Marmousi model. The correct velocity model is shown in Figure 7(a). The source locations are evenly distributed on the surface from 1.0 km to 7.0 km at a spacing of 0.1 km. The receiver arrays are fixed for all the shots and span entirely the surface at a spacing of 0.01 km. The data are generated via finite-difference Born modeling, using a Ricker wavelet with peak frequency of 15 Hz. The data are then transformed into the frequency domain because both the migration and wavefield tomography operators are based on the frequency-domain downward continuation method. In this way, we avoid using the same operator for both the modeling and inversion procedures. The image, space-lag gathers, and angle gathers migrated with the true velocity model are shown in Figures 7(b)-7(d). As one might expect, reflections in the space-lag gathers are focused at zero offset, and events in the angle gathers are flat because the correct model is used for mi-
Figure 2. (a) The true velocity model. The images migrated with (b) the true velocity model, and (c) the initial velocity model.
Figure 3. The space-lag CIGs constructed for the vertical reflector at \( x=1 \) km migrated with (a) the correct velocity, (b) the initial velocity, and CIGs at constructed for the horizontal reflector \( x=2.5 \) km migrated with (c) the correct velocity, (d) the initial velocity.
Figure 4. The CIPs at x=1 km, z=0.8 km migrated with (a) the correct velocity, (b) the initial velocity, compared with Figures 3(a) and 3(b). CIPs at x=1.7 km, z=2.5 km migrated with (c) the correct velocity, (d) the initial velocity, compared with Figures 3(c) and 3(d).
Figure 5. (a) The true velocity variation which is the target of inversion. The gradient constructed from (b) CIPs on the horizontal reflector only, (c) CIPs on the vertical reflector only, (d) CIPs on both the horizontal and vertical reflectors.
The initial model used for the inversion, Figure 8(a), is a highly smoothed version of the true model. This model resembles the results one can obtain from conventional ray-based reflection tomography. Figures 8(b)-8(d) show the image, space-lag gathers, and angle-domain CIGs migrated with the initial model. Since the initial model is highly smoothed, it lacks the necessary components required by the migration to produce an accurate result. Thus, the migrated image exhibits severe distortions in the reservoir region around \( x = 5 \text{ km}, \ z = 2.5 \text{ km} \). The reduced image quality can be further confirmed by the defocused energy in the space-lag gather and the residual moveout in the angle gathers, as shown in Figures 8(c) and 8(d). In our example, the angle gathers are used only for quality control, rather than for velocity model building.

We initiate our velocity analysis process by selecting common-image-point gather locations at which we construct extended images necessary for velocity analysis. These locations are selected using the automatic picking algorithm developed by Cullison and Sava (2011). Figure 8(b) shows the picked locations overlain on the image. They follow the coherent structure of the image, and tend to be randomly positioned where the reflections are less coherent. Figure 9 shows the weighting function we apply to the input gathers in order to compensate for the uneven illumination in the subsurface. Light colors indicate low values of the weights applied in the poor illumination areas, and dark colors indicate high value of the weights applied in the good illumination areas. This weighting function is included in the objective function to speed up the convergence of the inversion. Figures 10(a) and 10(b) show the inverted model after 20 nonlinear iterations and the corresponding migrated image. From the result, it is apparent that the updated velocity model significantly improves the imaging quality as the reservoir area is better focused and more coherent in the migrated image. In addition, the reflections are focused in space-lag gathers and flatter in angle gathers, as shown in Figures 10(c) and 10(d), as compared with those in Figures 8(c) and 8(d). These also indicate the improvement on the image due to the velocity update.

4 DISCUSSION

The synthetic examples demonstrate the successful velocity model updates produced by our approach using CIPs. In the first example, we show that CIPs do not bias their sensitivity toward any particular direction compared to the more conventional CIGs which sample image points more densely in the vertical direction than lateral direction. From the gradient computation, we notice that the CIPs located along nearly horizontal reflectors provide higher resolution laterally, whereas the CIPs located along nearly vertical reflectors provide higher resolution vertically.

In the second synthetic example, only the horizontal space lag and time lag are involved in CIPs computation. The vertical space lag is not required because there are no steeply dipping reflectors in the model. We thus significantly reduce...
Figure 7. (a) The true model used to generate the data. (b) The migrated image, (c) the space-lag gathers, and (d) the angle-domain gathers obtained using the true model.
Figure 8. (a) The initial model used in the velocity inversion. (b) The migrated image overlain with the CIPs location, (c) the space-lag gathers, and (d) the angle-domain gathers obtained using the initial model.
Figure 9. The weighting function based on the subsurface illumination.

the cost for computing and storing the gathers for velocity analysis. After running the inversion, the main features of the model are resolved and they help improving the quality of the image. The improvements can be directly observed from the more coherent image, more focused reflections in space-lag gathers, and flatter events in angle gathers. The results demonstrate our image-domain wavefield tomography based on CIPs can achieve similar results to conventional methods but with reduced computational effort. This is especially crucial for velocity analysis in large-scale 3D applications.

5 CONCLUSIONS

We demonstrate a wavefield-based velocity model building method implemented in the image domain. The procedure optimizes the velocity model by minimizing the image incoherency caused by model errors. The objective function is particularly designed for common-image-point gathers constructed locally on the reflection event. The penalty operator used in the objective function is aimed at improving the focusing of the reflections in the gathers.

The two synthetic examples demonstrate the main advantages of using CIPs over conventional subsurface-offset common-image gathers in the wave-equation tomographic approach. First, the CIPs avoid the bias toward horizontal reflection events and thus are more robust in analyzing velocity information for steeply dipping structure in the complex geologic regions. Second, CIPs significantly reduce the cost for computing and storing the extended images compared with more conventional common-image gathers while producing reliable model updates. This is mainly attributed to the optimized sampling of CIPs in the subsurface as only the significant reflections are analyzed to provide information for velocity update.

6 ACKNOWLEDGMENTS

We acknowledge the support of the sponsors of the Center for Wave Phenomena at Colorado School of Mines. The reproducible numeric examples in this paper use the Madagascar open-source software package freely available from http://www.reproducibility.org. This research was supported in part by the Golden Energy Computing Organization at the Colorado School of Mines using resources acquired with financial assistance from the National Science Foundation and the National Renewable Energy Laboratory.

REFERENCES

Mulder, W. A., 2008, Automatic velocity analysis with the two-way wave equation: Presented at the 70th EAGE Conference and Exhibition, Extended Abstracts.
Figure 10. (a) The updated model after 20 iterations of inversion using CIPs. (b) The migrated image overlain with the CIPs locations, (c) the space-lag gathers, and (d) the angle-domain gathers obtained using the updated model.


———, 2008b, Comparisons for waveform inversion, time domain or frequency domain?: 78th Annual International Meeting, SEG, Expanded Abstracts, 1890–1894.

Illumination compensation for image-domain wavefield tomography

Tongning Yang*, Jeffrey Shragge†, Paul Sava*

* Center for Wave Phenomena, Colorado School of Mines
† School of Earth and Environment, University of Western Australia

ABSTRACT

Wavefield tomography represents a family of velocity model building techniques based on full waveforms as the input and seismic wavefields as the information carrier. When these techniques are implemented in the image domain and use seismic images as the input, they are referred to as image-domain wavefield tomography. The objective function for image-domain approach is designed to optimize the coherency of reflections in extended common-image gathers. The function applies a penalty operator to the gathers, thus highlighting image inaccuracies due to the velocity model error. Minimizing the objective function optimizes the model and improves the image quality, by making use of the gradient of the objective function computed using the adjoint-state method. Uneven illumination is a common problem for complex geological regions, such as sub-salt, or the consequence of incomplete data. Imbalanced illumination not only creates shadow zone for migrated images, but also results in defocusing in common-image gathers even when the migration velocity model is correct. This additional defocusing violates the wavefield tomography assumption stating that the migrated images are perfectly focused in the case of the correct model. Therefore, defocusing rising from illumination mixes with defocusing rising from the model errors and degrades the model reconstruction. We address this problem by incorporating the illumination effects into the penalty operator such that only the defocusing by model errors is used for model construction. This method improves the robustness and effectiveness of wavefield tomography applied in the areas characterized by poor illumination. Our synthetic examples demonstrate that velocity models are more accurately reconstructed by our method using the illumination compensation, leading to more coherent and better focused subsurface images than those in the conventional approach without illumination compensation.

Key words: wavefield tomography, illumination compensation, extended images, adjoint state method

1 INTRODUCTION

Building an accurate and reliable velocity model remains one of the biggest challenges in current seismic imaging practice. In regions characterized by complex subsurface structure, prestack wave-equation depth migration, (e.g., one-way wave-equation migration or reverse-time migration), is a powerful tool for accurately imaging the earth’s interior (Gray et al., 2001; Etgen et al., 2009). The widespread use of these advanced imaging techniques drives the need for high-quality velocity models because these migration methods are very sensitive to model errors (Symes, 2008; Woodward et al., 2008; Virieux and Operto, 2009).

Wavefield tomography represents a family of techniques for velocity model building using seismic wavefields (Tarantola, 1984; Woodward, 1992; Pratt, 1999; Sirgue and Pratt, 2004; Plessix, 2006; Vigh and Starr, 2008; Plessix, 2009; Symes, 2009). The core of wavefield tomography is using a wave equation (typically constant density acoustic) to simulate wavefields as the information carrier. Wavefield tomog-
raphy is usually implemented in the data domain by adjusting the velocity model such that simulated and recorded data match (Tarantola, 1984; Pratt, 1999). This match is based on the strong assumption that the wave equation used for simulation is consistent with the physics of the earth. However, this is unlikely to be the case when the earth is characterized by strong (poro)elasticity. Significant effort is often directed toward removing the components of the recorded data that are inconsistent with the assumptions used.

Wavefield tomography can also be implemented in the image domain rather than in the data domain. Instead of minimizing the data misfit, the techniques in this category update the velocity model by optimizing the image quality, which is the cross-correlation of wavefields extrapolated from the source and receiver. The image quality is optimized when the data are migrated with the correct velocity model, as stated by the semblance principle (Al-Yahya, 1989; Yilmaz, 2001). The common idea is to optimize the coherence of reflection events in common-image gathers (CIGs) via velocity model updating. Since images are obtained using full seismograms, and velocity estimation also employs seismic wavefields as the information carrier, these techniques can be regarded as a particular type of wavefield tomography, and we refer to them as image-domain wavefield tomography. Unlike traditional ray-based reflection tomography methods, image-domain wavefield tomography uses band-limited wavefields in the optimization procedure. Thus, this technique is capable of handling complicated wave propagation phenomena such as multi-pathing in the subsurface. In addition, the band-limited character of the wave-equation engine more accurately approximates wave propagation in the subsurface and produces more reliable velocity updates.

Differential semblance optimization (DSO) is one realization of image-domain wavefield tomography. The essence of the method is to minimize the difference of the same reflectivity between neighboring offsets or angles. Symes and Carazzone (1991) propose a criterion for measuring coherency from offset gathers and establish the theoretic foundation for DSO. The concept is then generalized to space-lag (subsurface-offset) and angle-domain gathers (Shen and Calandra, 2005; Shen and Symes, 2008). Space-lag gathers (Rickett and Sava, 2002; Shen and Calandra, 2005) and angle-domain gathers (Sava and Fomel, 2003; Biondi and Symes, 2004) are two popular choices among various types of gathers used for velocity analysis. These gathers are obtained by wave-equation migration and are free of artifacts usually found in conventional offset gathers obtained by Kirchhoff migration, and thus they are suitable for applications in complex earth models (Stolk and Symes, 2004).

DSO implemented using space-lag gathers constructs a penalty operator which annihilates the energy at zero lag and enhances the energy at nonzero lags (Shen et al., 2003). This construction assumes that migrated images are perfectly focused at zero lag when the model is correct. If the model is incorrect, reflections in the gathers are defocused and the reflection energy spreads to nonzero lags. As a result, any energy left after applying the penalty is attributed to the result of model errors. However, this assumption is violated in practice when the subsurface illumination is uneven. Uneven illumination introduces additional defocusing such that images are not perfectly focused even if the velocity is correct, and it usually results from incomplete surface recorded data or from complex subsurface structure. Incomplete data cause loss of signal at some reflection angles and thus degrade the image. Nemeth et al. (1999) show that least-squared migration method can compensate the poor image quality due to the data deficiency. This method, however, is costly because of it requires many iterations to converge (Shen et al., 2011). In complex subsurface regions, such as sub-salt, uneven illumination is a general problem and it deteriorates the quality of imaging and velocity model building (Leveille et al., 2011). Gherasim et al. (2010) and Shen et al. (2011) show that the quality of migrated images can be optimized by illumination-based weighting generated from a demigration/remigration procedure. Tang and Biondi (2011) compute the diagonal of the Hessian matrix for the migration operator and use it for illumination compensation of the image before velocity analysis. These approaches effectively improve the quality of subsurface images, especially the balance of the amplitude. Nonetheless, they do not investigate the negative impact of illumination on the velocity model building. Furthermore, the misleading velocity updates due to uneven illumination remains an unsolved problem.

In this paper we address the problem of uneven illumination associated with image-domain wavefield tomography. We first review the theory for wavefield tomography in the image-domain including the formulation of the objective function and the gradient calculation. Next we explain how the uneven illumination affects the focusing of space-lag gathers and breaks down the assumption for DSO. We then propose a solution to this problem by using the illumination information in the construction of the penalty operator which is an integral part of the objective function. We illustrate our method with two synthetic examples representing two different types of illumination problems.

2 THEORY

In image-domain wavefield tomography using space-lag extended images (subsurface-offset CIGs), we formulate an objective function and compute its gradient using the adjoint-state method (Plessix, 2006; Symes, 2009). We discuss this method in detail and then analyze the influence of the illumination on gather focusing and how it impacts wavefield tomography. We conclude this section by introducing our solution to the illumination problem for wavefield tomography.

For simplicity, we discuss the methodology in the frequency-domain rather than in the time-domain although the latter is completely equivalent and analogous. We formulate the inverse problem by first defining the state variables, through which the objective function is related to the model parameter. The state variables for our problem are the source and receiver wavefields $u_s$ and $u_r$ obtained by solving the fol-
Wavefield tomography with illumination compensation

lowing acoustic wave equation:
\[
\begin{bmatrix}
\mathcal{L}(x, \omega, m) & 0 \\
0 & \mathcal{L}^*(x, \omega, m)
\end{bmatrix}
\begin{bmatrix}
\bar{u}_s(j, x, \omega) \\
\bar{u}_r(j, x, \omega)
\end{bmatrix}
= \begin{bmatrix}
\bar{f}_s(j, x, \omega) \\
\bar{f}_r(j, x, \omega)
\end{bmatrix},
\tag{1}
\]
where \(f_s\) is a point or plane source, \(f_r\) are the record data, \(j = 1, 2, \ldots, N_s\) where \(N_s\) is the number of shots, \(\omega\) is the angular frequency, and \(x\) are the coordinates \((x, y, z)\). The wave operator \(\mathcal{L}\) and its adjoint \(\mathcal{L}^*\) propagate the wavefields forward and backward in time respectively using a two-way wave equation. Thus, \(\mathcal{L}\) is formulated as
\[
\mathcal{L} = -\omega^2 m - \Delta,
\tag{2}
\]
where \(\Delta\) is the Laplace operator, and \(m\) represents the model (slowness squared).

In the second step of the adjoint-state method, we construct the objective function from which we derive the adjoint sources used to model the adjoint-state variables required by the gradient computation. The objective function for image-domain wavefield tomography is defined using the semblance principle (Yilmaz, 2001) and measures the image incoherency caused by the model errors. Therefore, the inversion simultaneously reconstructs the model and improves the image quality by minimizing the objective function.

The objective function based on space-lag CIGs is
\[
\mathcal{H}_\lambda = \frac{1}{2} \left\| K_I(x) P(\lambda) r(x, \lambda) \right\|^2_{\lambda, \lambda},
\tag{3}
\]
and
\[
r(x, \lambda) = \sum_j \sum_\omega \bar{u}_s(j, x - \lambda, \omega) u_r(j, x + \lambda, \omega)
= \sum_j \sum_\omega T(-\lambda) \bar{u}_s(j, x, \lambda^T) u_r(j, x, \lambda),
\tag{4}
\]
where the overline represents complex conjugate. The operator \(T^\prime\) represents the space shift applied to the wavefields and is defined by
\[
T(\lambda) u(j, x, \omega) = u(j, x + \lambda, \omega).
\tag{5}
\]
The mask operator \(K_I(x)\) limits the construction of the gather to select locations in the subsurface. The penalty operator \(P(\lambda)\) annihilates the focused energy at zero lag and highlights the energy of residual moveout at nonzero lag (Shen and Symes, 2008):
\[
P(\lambda) = |\lambda|,
\tag{6}
\]
In practice, it is common to restrict the space-lags in the horizontal directions only, i.e. \(\lambda = \{\lambda_x, \lambda_y, 0\}\). The objective function \(\mathcal{H}_\lambda\) is minimized when the reflections focus at zero lag, which is an indication of the correct velocity model.

The adjoint sources are computed as the derivatives of the objective function \(\mathcal{H}_\lambda\) shown in equation 3 with respect to the state variables \(u_s\) and \(u_r\):
\[
\begin{bmatrix}
g_s(j, x, \omega) \\
g_r(j, x, \omega)
\end{bmatrix} =
\begin{bmatrix}
\sum_\lambda T(\lambda) P(\lambda) K_I(x) K_I^T(x) P(\lambda) r(x, \lambda) \mathcal{H}_\lambda u_r(j, x, \omega) \\
\sum_{\lambda} T(-\lambda) P(\lambda) K_I(x) K_I^T(x) P(\lambda) r(x, \lambda) \mathcal{H}_\lambda u_s(j, x, \omega)
\end{bmatrix}.
\tag{7}
\]
The adjoint state variables \(a_s\) and \(a_r\) are the wavefields obtained by backward and forward modeling, respectively, using the corresponding adjoint sources defined in equation 7:
\[
\begin{bmatrix}
a_s(j, x, \omega) \\
a_r(j, x, \omega)
\end{bmatrix} =
\begin{bmatrix}
g_s(j, x, \omega) \\
g_r(j, x, \omega)
\end{bmatrix},
\tag{8}
\]
where \(\mathcal{L}\) and \(\mathcal{L}^*\) are the same wave propagation operators used in equation 1.

The last step of the gradient computation is simply the correlation between state variables and adjoint state variables:
\[
\frac{\partial \mathcal{H}_\lambda}{\partial m} = \sum_j \sum_\omega \frac{\partial \mathcal{L}}{\partial m} \begin{bmatrix}
u_s(j, x, \omega) \ a_s(j, x, \omega) + u_r(j, x, \omega) a_r(j, x, \omega)
\end{bmatrix},
\tag{9}
\]
where \(\frac{\partial \mathcal{L}}{\partial m}\) is the partial derivative of the wave propagation operator with respect to the model parameter. Using the definition of \(\mathcal{L}\) in equation 2, we find that \(\frac{\partial \mathcal{L}}{\partial m} = -\omega^2\). From equation 9 one may notice that the gradient for image-domain wavefield tomography consists of two correlations because we define both the source and receiver wavefields as the state variables.

The derivation above shows the construction of the image-domain wavefield tomography objective function and its gradient. Given these two components, the solution to the inverse problem is found by minimizing the objective function using non-linear gradient-based iterative methods (Knyazev and Lashuk, 2007). In each iteration, the gradient is computed and the model update is calculated by a line search in the steepest descent or conjugate gradient directions.

To analyze the illumination effects on focusing, we first illustrate the focusing mechanism for reflections in space-lag gathers. Yang and Sava (2010) derive the analytic formula for the reflection moveout in the extended images. Assuming constant local velocity, a reflection in space-lag gathers obtained by migrating one shot experiment is a straight line in 2D and a plane in 3D:
\[
z(\lambda) = d_0 - \frac{\tan \theta(q \cdot \lambda)}{n_z}.
\tag{10}
\]
Here \(d_0\) represents the depth of the reflection corresponding to the chosen CIGs location, \(\lambda\) represents the horizontal space lag \(\lambda = \{\lambda_x, \lambda_y, 0\}\), \(n_z\) is the vertical component of the unit vector normal to the reflection plane, \(q\) is the unit vector parallel to the reflection plane, and \(\theta\) is the reflection angle. When we stack the gathers obtained with a correct model from all available shots, the straight lines or planes corresponding to reflections from different shots are superimposed and interfere to form a focused point at zero lag (Yang and Sava, 2010). Good
interference of such events, however, occurs only if the reflector point is well illuminated by the experiments from the surface. In other words, the surface shot coverage must be large enough and regularly distributed. Also, the subsurface illumination must be even so that the reflector is illuminated on a sufficient range of reflection angles from the surface. If either of these conditions is not satisfied, the reflection energy does not interfere perfectly and the gather defocus, even if the correct model is used for the imaging.

We conclude that the defocusing of reflections in the space-lag gathers may result from velocity errors, as well as from the imperfect stacking due to uneven illumination. In general, these two different kinds of defocusing are indistinguishable to the velocity analysis procedure. The penalty operator in equation 6 emphasizes all energy away from zero lag and includes both the defocusing caused by the velocity error and by the uneven illumination. Thus, the penalty operator leads to a residual that is much larger than what would be expected for a given error in the model. Penalizing the defocusing due to the illumination misleads the inversion and results in overcorrection and artificial updates.

To alleviate the negative influence of uneven illumination, we need to include the illumination information in the tomographic procedure. One approach uses the illumination information as a weighting function to preconditioning the gathers. The gathers in poor illumination areas are down-weighted as the defocusing information is less reliable than the gathers in good illumination areas. This approach stabilizes the inversion but decreases the accuracy of the results as the useful velocity information in poor illumination areas is ignored. The poor illumination area, however, is exactly the place where we want to make use of all available information for the model building. To overcome the illumination problem and preserve the useful information in the tomography, we introduce a new illumination-based penalty operator to replace the conventional DSO penalty operator. The new operator is constructed such that it only emphasizes the defocusing caused by the velocity error and ignores the defocusing caused by uneven illumination. To achieve this goal, we analyze the image defocusing due to uneven illumination by applying illumination analysis.

Illumination analysis in the framework of wave-equation migration is formulated using the solution to migration deconvolution problems (Yu and Schuster, 2003). Migration deconvolution first establishes a linear relationship between a reflectivity distribution \( \tilde{r} \) and seismic data \( d \):

\[
\mathcal{M} \tilde{r} (\mathbf{x}) = d ,
\]

where \( \mathcal{M} \) represents a forward Born modeling operator which is linear with respect to the reflectivity. A migrated image is obtained by applying the adjoint of the modeling operator \( \mathcal{M}^* \) to the data,

\[
\mathcal{M}^* d = \mathcal{M}^* \mathcal{M} \tilde{r} (\mathbf{x}) = r (\mathbf{x}) ,
\]

where \( r \) is a migrated image. Note that the migrated image is the result of blurring the reflectivity \( \tilde{r} \) by \( \mathcal{M}^* \mathcal{M} \), which is the Hessian (second-order derivative of the operator with respect to the model) for the operator \( \mathcal{M} \). Thus, the reflectivity can be computed from the migrated image by

\[
\tilde{r} (\mathbf{x}) = (\mathcal{M}^* \mathcal{M})^{-1} r (\mathbf{x}) ,
\]

where \( (\mathcal{M}^* \mathcal{M})^{-1} \) includes the subsurface illumination information associated with the velocity structure and acquisition geometry. In practice, the full \( (\mathcal{M}^* \mathcal{M})^{-1} \) matrix is too costly to construct, but we can evaluate its impact by applying a cascade of demigration and migration \( \mathcal{M}^* \mathcal{M} \) to a reference image:

\[
r_e (\mathbf{x}) = \mathcal{M}^* \mathcal{M} r (\mathbf{x}) .
\]

The resulting image \( r_e \) approximates the diagonal elements of the Hessian as the illumination effects, and can be used as a weight for illumination compensation of migrated image (Guitton, 2004; Gherasim et al., 2010; Tang and Biondi, 2011).

Equation 14 only computes the illumination effects distributed at image points. In our problem, we are concerned with the defocusing caused by the illumination in the space-lag gathers, and need to evaluate the illumination effects at image points and along their space-lag extension. Using the concept of extended modeling (Symes, 2008), we can generalize equation 12 to the extended space \( \mathbf{x} - \lambda \):

\[
r_e (\mathbf{x}, \lambda) = \mathcal{M}^* \mathcal{M} r (\mathbf{x}, \lambda) ,
\]

where \( r_e (\mathbf{x}, \lambda) \) are space-lag gathers containing reference image at zero space lag, and \( r_e \) are the output gathers containing defocusing associated with illumination effects. Such defocusing is the consequence of uneven illumination and should not be penalized by the penalty operator in the velocity updating process. Therefore, an illumination-based penalty operator can be constructed as

\[
P (\mathbf{x}, \lambda) = \frac{1}{E[r_e (\mathbf{x}, \lambda)] + \epsilon} ,
\]

where \( E \) represents envelope and \( \epsilon \) is a damping factor used to stabilize the division. By definition, this penalty operator has low values in the area of defocusing due to uneven illumination and high values in the rest. Thus, this operator is consistent with our idea of avoiding penalty to reflection energy irrelevant to velocity errors, e.g., the artifacts caused by illumination. Replacing the conventional penalty in equation 6 with the one in equation 16 is the basis for our illumination compensated image-domain wavefield tomography. Note that the DSO penalty operator is a special case of our new penalty operator and corresponds to the case of perfect subsurface illumination and wide-band data.

3 EXAMPLES

In this section, we use two synthetic examples to illustrate our illumination compensated image-domain wavefield tomography. In the first example, we use incomplete data to simulate illumination problems due to the acquisition. In the second example, we use the Sigsbee model to test our method in regions of complex geology with poor subsurface illumination caused by irregular salt.
Figure 1. (a) The true model used to generate the data, (b) the migrated image, and (c) the angle-domain gathers obtained using the true model.
Figure 2. Two shot gathers at (a) 0.5 km and (b) 2.0 km showing a gap of 0.6 km in the acquisition surface. The gap simulate an obstacle which prevents data acquisition, e.g., a drilling platform.
Figure 3. (a) The initial constant model, (b) the migrated image, and (c) the angle-domain gathers obtained using the initial model.
Figure 4. (a) The conventional DSO penalty operator. (c) The gathers obtained with demigration/migration showing the illumination effects. (d) The illumination-based penalty operator constructed from the gathers in Figure 4(c). The light areas cover the defocusing due to the illumination.
Figure 5. (a) The reconstructed model using conventional DSO penalty, (b) the migrated image, and (c) the angle-domain gathers obtained using the reconstructed model.
Figure 6. (a) The reconstructed model using the illumination-based penalty, (b) the migrated image, and (c) the angle-domain gathers obtained using the reconstructed model.
Wavefield tomography with illumination compensation

The velocity model for the first example is shown in Figure 1(a). For five horizontal interfaces simulated as density contrasts, we generate the data at receivers distributed along the surface. Two shot gathers are shown in Figures 2(a) and 2(b). The data are truncated from 2.2 – 2.8 km to simulate an acquisition gap. The initial model is constant (Figures 3(a)). The migrated image and angle-domain gathers for the true and initial models are shown in Figures 1(b)-1(c) and Figures 3(b)-3(c), respectively. Note that the angle gather is displayed at selected locations corresponding to the vertical bars overlain in Figures 1(b) and 3(b). The migrated image for the initial model shows defocusing and crossing events caused by the incorrect model. The illumination gap due to the missing data and the residual moveout caused by the wrong velocity can be observed on the angle gather shown in Figure 3(c).

For comparison with our method, we run the inversion using conventional DSO penalty operator. Figure 9(a) plots the penalty operators at the same selected locations shown in Figure 1(b). The actual spacing of the gathers and penalty operators are 0.2 km. Since the conventional penalty is laterally invariant, the figure consists of the same operators duplicated at different lateral position. The inverted model after 30 nonlinear iterations is plotted in Figure 5(a), and the corresponding migrated image and angle gather are shown in Figure 5(b) and 5(c). The reconstruction of the model is not satisfactory, especially for the anomaly under the acquisition gap. Also, the model contains many artifacts as the consequence of the incomplete data. As a result, the reduced quality of the image obtained with the inverted model is not surprising. Although the reflections on the left of the image are quite continuous and flat, those on the right, especially under the acquisition gap, are not flat and are even discontinuous. This result clearly shows the negative impact of the poor illumination on image-domain wavefield tomography.

To show the defocusing due to illumination effects in the gathers, we construct the space-lag gathers with the current image in Figure 3(b) at zero-lag as the reference image gather. We apply the demigration/migration workflow from equation 15 to the reference gather. The result, as shown in Figure 4(b), characterizes the illumination effects given the velocity model and acquisition setup. Most reflection energy is focused, thus indicating good illumination, but we can still observe defocusing as the consequence of incomplete data in the area under the acquisition gap. The illumination-based penalty operator is constructed from the gathers in Figure 4(b) using equation 16, as plotted in Figure 4(c). We can observe that the areas in light color coincide with the focused energy at zero lag and defocusing away from zero lag in Figure 4(b). Thus, the penalty operator does not emphasize the defocusing due to the uneven illumination and highlights only the defocusing due to velocity errors.

Using the new penalty operator, we update the model under the same conditions as in the example using DSO. Figure 6(a). Compared with the result in Figure 5(a), the result obtained with the new penalty is cleaner and contains fewer artifacts. Also, the anomaly under the acquisition gap is more accurately reconstructed and closer to the true model. Because of the improved model, the images are also improved, as seen in Figure 6(b). The reflections under the acquisition gap are more continuous and flat than the image in Figure 5(b). In addition, one can observe from the angle gather in Figure 6(c) that more events appear in the area under the acquisition gap and overall reflections are flatter, which indicates an improved signal-to-noise ratio arising from the more accurate reconstructed model.

We also apply our method to the Sigsbee 2A model (Paffenholz et al., 2002), and concentrate on the subsalt region. The target area ranges from $x = 6.5 - 20$ km, and from $z = 4.5 - 9$ km. The model, migrated image, and angle-domain gather for correct and initial models are shown in Figures 7(a)-7(c) and Figures 8(a)-8(c), respectively. The angle gather is displayed at selected locations corresponding to the vertical bars overlain in Figures 7(b) and 8(b). The actual spacing of the gatherings and penalty operators are 0.45 km. Note that the reflections in the angle gather appear only at positive angles, as the data are simulated for towed streamers and the subsurface is illuminated from one side only. Just as for the previous example, we run the inversion using both the conventional DSO penalty and the illumination-based penalty operators. The DSO penalty operator is shown in Figure 9(a). For the illumination-based operator, we first generate gathers containing defocusing due to illumination (Figure 9(b)), and then we construct the penalty operator using equation 16 as shown in Figure 9(c). From the gathers characterizing the illumination effects, we can observe the significant defocusing in the subsalt area as the salt distorts the wavefields used for imaging and causes the poor illumination in this area.

We run both inversions for 10 iterations, and obtain the reconstructed model, migrated image, and angle-domain gathers shown in Figures 10(a)-10(c) and Figures 11(a)-11(c), respectively. The figures show that we update the models in the correct direction in both cases and that the reconstructed models are closer to the true model than the starting model. We find, however, that the model obtained using the illumination-based penalty is closer to the true model than the model obtained using DSO penalty. The DSO model is not sufficiently updated and is too slow. This is because the severe defocusing due to the salt biases the inversion when we do not take into accounts the uneven illumination for tomography. The comparison of the images also suggests that the inversion using the illumination-based penalty is superior to the inversion using DSO penalty. Both the images are improved due to the updated model, as illustrated, for example, by that the diffractions distributed at $z = 7.6$ km focus, and the faults located between $x = 14.0$ km, $z = 6.0$ km and $x = 16.0$ km, $z = 9.0$ km are more visible in the images. If we concentrate on the bottom reflector (around 9 km), we can distinguish the extent of the improvements on the image quality for both inversions. The bottom reflector is corrected to the right depth for inversion using the illumination-based penalty, while the bottom reflector for inversion using DSO penalty is still away from the right depth and not as flat as the reflector in Figure 11(b). Figures 12(a)-12(d) compare the angle gathers at $x = 10.2$ km for the correct, initial, and reconstructed models using DSO
Figure 7. (a) The true model in the target area of the Sigsbee model. (b) The migrated image, and (c) the angle-domain gathers obtained using the true model.
Figure 8. (a) The initial model obtained by scaling the true model. (b) The migrated image, and (c) the angle-domain gathers obtained using the initial model.
Figure 9. (a) The conventional DSO penalty operator. (b) The gathers obtained with demigration/migration showing the illumination effects. (c) The illumination-based penalty operator constructed from the gathers in Figure 9(b). The light areas cover the defocusing due to the illumination.
Figure 10. (a) The reconstructed model from inversion using the DSO penalty. (b) The migrated image, and (c) the angle-domain gathers obtained using the reconstructed model.
Figure 11. (a) The reconstructed model from inversion using the illumination-based penalty. (b) The migrated image, and (c) the angle-domain gathers obtained using the reconstructed model.
Wavefield tomography with illumination compensation

Figure 12. Angle-domain gathers at \( x = 10.2 \) km for (a) the correct model, (b) the initial model, (c) the reconstructed model using DSO penalty, and (d) the reconstructed model using the illumination-based penalty.

and illumination-based penalties. The gathers for both reconstructed models show flatter reflections, indicates that reconstructed models are more accurate than the initial model. We can, nonetheless, observe that the reflections in Figure 12(d) are flatter than those in Figure 12(c), and conclude that the reconstructed model using the illumination-based penalty is more accurate.

4 DISCUSSION

Uneven illumination is often a challenge faced by the exploration activities in complex subsurface environments, particularly in sub-salt areas. Furthermore, imperfect acquisition can also cause illumination problems. Essentially, the core of the illumination problem is that the reflections from various angles cannot be observed on the surface, either because of the complexity of the subsurface or because of limited/partial acquisition on the surface. In both situations, the effectiveness of image-domain wavefield tomography deteriorates as the imbalanced illumination creates defocusing in space-lag gathers regardless of the accuracy of the velocity model. Minimizing such defocusing through wavefield tomography generates incorrect updates and artifacts in the result. Therefore, the defocusing caused by the uneven illumination must be excluded from the model building process. This is done by replacing the DSO penalty operator with a modified penalty operator, which is constructed based on the illumination information. This penalty operator differentiates between the defocusing caused by uneven illumination and that due to model errors and therefore does not penalize the defocusing due to the illumination. During our iterative inversion, we update penalty operator at each iteration, as the velocity model is changing during the iterations and the subsurface illumination information needs to be re-evaluated.

In the theory section, we discuss the formulation of image-domain wavefield tomography and our solution to illumination problems based on a two-way wave equation. The formulation, however, applies well when a one-way wave equation is used. The construction of the illumination-based penalty operator is similar, except for the wave equation used in all modeling and migrations.

5 CONCLUSIONS

We demonstrate an illumination compensation strategy for wavefield tomography in the image domain. The idea is to measure the illumination effects on space-lag extended images, and replace the conventional DSO penalty operator with another one that compensates for illumination. This approach isolates the defocusing caused by the illumination such that image-domain wavefield tomography minimizes only the defocusing relevant to the velocity error. Synthetic examples demonstrate the negative effects of uneven illumination on the reconstructed model and show the improvements on both the inversion results and on the migrated images after the illumination information is included in the penalty operator. Our approach enhances the robustness and effectiveness of wavefield tomography in the model building process when the surface data are incomplete or when the subsurface illumination is uneven due to complex geologic structures such as salt.

6 ACKNOWLEDGMENTS

We acknowledge the support of the sponsors of the Center for Wave Phenomena at Colorado School of Mines. The reproducible numeric examples in this paper use the Madagascar open-source software package freely available from http://www.reproducibility.org. This research was supported in part by the Golden Energy Computing Organization at the Colorado School of Mines using resources acquired with financial
assistance from the National Science Foundation and the National Renewable Energy Laboratory.

REFERENCES


Understanding the reverse time migration backscattering: Noise or signal?

Esteban Díaz and Paul Sava

Center for Wave Phenomena, Colorado School of Mines, Golden CO 80401

ABSTRACT
Reverse time migration (RTM) backscattered events are produced by the cross-correlation between waves reflected from sharp interfaces (e.g. the top of salt bodies). These events, along with head waves and diving waves, produce the so-called RTM artifacts, which are visible as low wavenumber energy on migrated images. Commonly, these events are seen as a drawback for the RTM method because they obstruct the image of the geologic structure, which is the real objective for the process. Many strategies have been developed to filter out the artifacts from the conventional image. However, these events contain information that can be used to analyze kinematic synchronization between source and receiver wavefields reconstructed in the subsurface. Numeric and theoretical analysis indicate the sensitivity of the backscattered energy to velocity accuracy: an accurate velocity model maximizes the backscattered artifacts. The analysis of RTM extended images can be used as a quality control tool and as input to velocity analysis to constrain salt models and sediment velocity.

Key words: RTM backscattering, wavefield decomposition, migration velocity analysis

1 INTRODUCTION
Reverse time migration (RTM) is not a new imaging technique (Baysal et al., 1983; Whitmore, 1983; McMechan, 1983). However, it was not until the late 1990s, and mainly the 2000s that computational advances allowed the geophysical community to use this technology for exploratory 3D surveys. In general, and especially in complex geological settings, RTM produces better images than other methods. Imaging methods such as Kirchhoff migration and one-way equation migration are based on approximate solutions to the wave equation. Kirchhoff migration, a high frequency asymptotic solution to the wave equation, becomes unstable for complex velocity models. This technique also fails to handle multipathing and typically creates the images based on a single travel-time arrival (e.g. most energetic or first arrival). Other methods based on approximations to the wave equation, such as phase shift migration (Gazdag, 1978), rely on a v(z) earth model and further approximations are needed to account for lateral variations (Gazdag and Sguazzero, 1984). In addition to earth model considerations, one-way wave equation migration propagates the wavefields in either the upward or the downward direction; this approximation becomes inexact when the waves propagate horizontally. Therefore, this technique fails to properly handle overturning waves and reflections from steep-dip structures. RTM’s propagation engine, a two-way wave equation, makes this imaging method robust and accurate because it honors the kinematics of the wave phenomena by allowing waves to propagate in all directions regardless of the velocity model or the direction of propagation. This method also takes into account, in a natural way, multipathing and reflections from steep dips.

A striking characteristic of RTM is the presence of low wavenumber events in the image that are uncorrelated with the geology. The two-way wave equation simulates scattered waves in all directions. Therefore, the imaging condition produces new events not observed in other imaging methods that correspond to the cross-correlation between diving waves, head waves and backscattered waves. The cross-correlation between the backscattered waves is more visible in presence of sharp boundaries (e.g. the top of salt) which produces strong events that mask the image of the earth reflectivity above the salt. The backscattered events are considered as noise and are normally filtered in order to get the image of earth reflectivity.

The seismic industry has dedicated effort and time developing algorithms and strategies to filter out the backscattered energy from the image. We can classify the filtering approaches in two general families: pre-imaging condition and post-imaging condition.

The pre-imaging condition family modify the wavefields (either by modeling or wavefield decomposition) in such a way that the backscattered events do not form during the imaging
process. One strategy in the pre-imaging condition category is wavefield decomposition (Liu et al., 2011; Fei et al., 2010). In this method, the source and receiver wavefield are decomposed in upgoing and downgoing directions. In the imaging step, we cross-correlate only the wavefields that propagate in opposite directions producing an image which corresponds to the geology. The cross-correlation between wavefields traveling in parallel directions is discarded because produce events that obstructs the geology. Other pre-imaging condition approaches are performed by modifying the wave equation to attenuate the reflections coming from sharp interfaces (Fletcher et al., 2005). A similar method applicable to post-stack migration uses impedance matching at sharp interfaces (Baysal et al., 1984).

In the post-imaging family, the artifacts are attenuated by filtering. These filtering approaches are considerably cheaper because they operate in the image space and not on the wavefields. A straightforward approach is to apply a Laplacian operator to the image (Youn and Zhou, 2001); this operator acts as a high pass filter and is effective because the backscattered events have a strong low wavenumber component. A second strategy is a signal/noise separation by least squares filtering. In this case the signal is defined as the reflectivity and the noise is the backscattered energy (Guitton et al., 2007). Finally, extended imaging conditions (Rickett and Sava, 2002; Sava and Fomel, 2006; Sava and Vasconcelos, 2011) provide information about the wavefield similarity for different space and/or time lags and can also be used to discriminate the backscattered energy. Kaelin and Carvajal (2011) take advantage of the way backscattered events appear in time-lag gathers. The backscattered events map toward zero time-lag when a correct velocity model is used for imaging, whereas the primary reflections map within a limited slope range constrained by the velocity model. This difference in slope allows us to design 2D filters that preserve events within the primaries reflections range and attenuate the backscattered energy.

In this report we analyze the information carried by the backscattered energy in the extended images. We show that the backscattered waves provide important information about the synchronization between the reconstructed wavefields in the subsurface, i.e. an image obtained with a correct velocity model shows maximum backscattered energy. The presence of backscattered energy in the image not only depends on the interpretation of the sharp interface but also on the velocity above it. We analyze the mapping patterns of the backscattered events in the extended images using wavefield decomposition approaches and conclude that backscattered energy is sensitive to the velocity model accuracy and therefore should be included as a source of information to migration velocity analysis (MVA). Counter to common practice, we assert that backscattering artifacts should be enhanced during RTM to constrain the velocity models, and they should only be removed in the last stage of imaging.

2 WAVE EQUATION IMAGING CONDITIONS

2.1 Conventional imaging condition

The conventional imaging condition (Claerbout, 1985) is a zero time lag cross-correlation between the source wavefield and the receiver wavefields:

$$ R(x) = \sum_{\text{shots}} \sum_{t} W_s(x,t) W_r(x,t), $$

which honors the single scattering or Born assumption. Under this assumption the transmitted source wavefield generates secondary waves as it interacts with the medium discontinuities. These secondary waves propagate in space and are recorded at the surface. This assumption means that both the source and receiver wavefields carry only transmitted energy through interfaces between layers with different elastic properties.

A wavefield extrapolated with RTM could show, depending on the complexity of the geology, waves traveling in both upward and downward directions, such as diverging waves, head waves and backscattered waves. The interaction between these waves contained in the source and receiver wavefields generates new events in the image which are commonly referred to as artifacts because they do not follow the geology (i.e. earth reflectivity), which is the objective of the imaging process. The correlation between forward and backscattered waves is particularly strong when sharp boundaries are present in the velocity model (e.g. salt bodies).

If a sharp boundary is present in the model, we can decompose the source wavefield into transmitted and reflected energy that originates at the sharp boundary:

$$ W_s(x,t) = W_s^t(x,t) + W_s^r(x,t), $$

where the superscripts $t$ and $r$ stand for transmitted and reflected energy, respectively.

The same idea can be applied to the receiver wavefield:

$$ W_r(x,t) = W_r^t(x,t) + W_r^r(x,t). $$

By taking advantage of the linearity of equation 1, we can split the conventional imaging condition as follows:

$$ R(x) = R^{tt}(x) + R^{rt}(x) + R^{tr}(x) + R^{rr}(x). $$

Here, the first superscript is associated with the source wavefield and the second is associated to the receiver wavefield. For example, $R^{tt}(x)$ is an image constructed with the transmitted source wavefield and the reflected receiver wavefield.

By analyzing the individual contributions to the image, we can better understand how the backscattered events are constructed in the image. This analysis is similar to the one of Fei et al. (2010) and Liu et al. (2011) whose objective is to filter out the non-geological portions of the image. Here, we approach the problem in a broader sense by attempting to un-
understand the physical meaning of the backscattered energy and its uses for velocity model building.

### 2.1.1 Backscattered events in the conventional image

In order to gain an understanding of the RTM backscattered events, we use a simple model with two-layers and strong velocity contrast. Figure 1(d) shows the image obtained with the conventional imaging condition for one shot at \( x = 5\, \text{km} \). This image has strong backscattered energy, indicated with letter “a”, above the reflector located at \( z = 1.5\, \text{km} \).

To better understand the origin of the backscattered artifacts, we illustrate the wavefields used for imaging our simple model. Figures 2(a), 2(e) and 2(i) show three different snapshots of the source wavefield. Likewise, Figures 2(b), 2(f) and 2(j) show the same snapshots for the receiver wavefield. Figures 2(c), 2(g) and 2(k) show the product between source and receiver wavefields for the same time snapshots. Finally, Figures 2(d), 2(h) and 2(l) show the accumulated image as a function of time (integration over time of the product between wavefields).

Figure 2(d) show the interaction between the transmitted source wavefield \( W_t^s \), shown in Figure 2(a), and the reflected receiver wavefield \( W_r^r \), shown in Figure 2(b). In this case, the reflected receiver wavefield travels in perfect synchronization with the transmitted source wavefield, therefore their product, shown in Figure 2(c), stacks coherently in the imaging process generating the \( R^{tr}(x) \) contribution to the image \( R(x) \). In the \( R^{rt}(x) \) image, the reflected receiver wavefield behaves as the transmitted source wavefield, which is the reason why the backscattered energy is imaged toward the source location. In the partial image at \( t = 0.275\, \text{s} \), shown in Figure 2(h), we see how we start building the reflector image. The reflected source wavefield, shown in Figure 2(e), generates new backscattered events corresponding to the \( R^{tr}(x) \) image. In the snapshot at \( t = 0.5\, \text{s} \), the reflector is completely imaged and for remaining time we only add backscattered energy corresponding to the \( R^{rt}(x) \) image. In this case the reflected source wavefield behaves as the receiver wavefield and its energy maps toward the receivers. We can see that after the imaging process is finished, Figure 1(d), that the backscattered energy is maximum near the critical angle range (where the reflected source and receiver wavefields have maximum energy).

Using wavefield decomposition allow us to isolate the individual contributions of equation 4. Figure 3(a) shows the cross-correlation between transmitted wavefields, producing an image due to the earth reflectivity. Figures 3(b) and 3(c) show the images \( R^{tr}(x) \) and \( R^{rt}(x) \) corresponding to the backscattered energy, which maps toward the source and the receivers, respectively. The image corresponding to the \( R^{rr}(x) \), shown in Figure 3(d), contains additional contribution to the reflectivity of the earth due to the cross-correlation between reflected wavefields. Fei et al. (2010) take advantage of this analysis to define an image free from backscattered energy as \( R(x) = R^{tt}(x) + R^{rr}(x) \). Here, we want to better understand the meaning and uses of the other two partial images \( R^{tr}(x) \) and \( R^{rt}(x) \).
Figure 2. Pictorial explanation of RTM imaging: rows 1, 2 and 3 correspond to three different snapshots at times $t_1 = 0.150s$, $t_2 = 0.275s$ and $t_3 = 0.500s$. Columns 1 to 4 correspond to the source wavefield, the receiver wavefield, the multiplication of the source and receiver wavefields, and the accumulated image over time, respectively.
2.2 Extended imaging condition

The extended imaging condition (Rickett and Sava, 2002; Sava and Fomel, 2006; Sava and Vasconcelos, 2011) is similar to the conventional imaging condition except the cross-correlation lags between source and receiver wavefield are preserved in the output:

$$R(x, \lambda, \tau) = \sum_{\text{shots}} \sum_{t} W_s(x - \lambda, t - \tau) W_r(x + \lambda, t + \tau).$$

Here $\lambda$ and $\tau$ represent the space-lags and time-lags, respectively, of the cross-correlation. The conventional image is a special case of the extended image $R(x) = R(x, 0, 0)$.

By using extended images, we can measure the accuracy of the velocity model by analyzing the moveout of the events (Yang and Sava, 2010), and we can perform transformations from the extended to the angle domain (Sava and Fomel, 2003, 2006; Sava and Vlad, 2011). The extended images provide a measurement of the similarity between the source and receiver wavefields along space and time, so we can exploit these images to analyze and better understand the RTM backscattered events.

In equation 5 we observe an increase in the dimensionality of the image, from 3 to 7 dimensions, if we decide to extend the image in all directions. It is common to perform the analysis of extended images at limited locations in order to make this methodology feasible for large datasets. For cost considerations we often use an extension for common image gathers (CIG), for instance the time-lag axis ($\tau$) or the space-lag axis ($\lambda$). We can also consider common image point gathers (CIP), where we fix an observation point $c = (x, y, z)$ and analyze the image as a function of extensions $\lambda, \tau$.

Figures 1(a) to 1(c) show a time-lag gather, a space-lag gather, and a common image point, respectively, which represent subsets at fixed surface positions (for CIGs) or fixed space positions (for CIPs). Despite the fact that our model has only one reflector, we can identify several events in the conventional and extended images. Letter “a” indicates backscattered events, letter “b” indicates the events produced by the cross-correlation of reflected wavefields, and letter “c” indicates the cross-correlation between transmitted wavefields.

In the presence of sharp velocity interfaces we can use the concept of equation 4, and construct four partial extended images:

$$R(x, \lambda, \tau) = R^{tt}(x, \lambda, \tau) + R^{rr}(x, \lambda, \tau) + R^{tr}(x, \lambda, \tau) + R^{rt}(x, \lambda, \tau).$$

2.2.1 Time-lag common image gathers

Using equation 6, we analyze the individual contributions for the time-lag gather shown in Figure 1(a). Figure 4(a) shows the image $R^{tt}(z, \tau)$ with a change in the slope of the events due to the abrupt velocity variation of the model. Above the reflector depth, the slope is controlled by the velocity of layer 1, whereas below the interface the slope is controlled by the velocity of layer 2. Figures 4(c) and 4(b) show the backscattered event contributions $R^{tt}(z, \tau)$ and $R^{rr}(z, \tau)$, respectively, which indicate that the backscattered event maps towards $\tau=0$ in the extended image. This means that we only get a contribution when we do not dislocate the wavefields by shifting them in time, thus reinforcing the idea of wavefield synchronization. Figure 4(d) shows the $R^{rr}(z, \tau)$ image; in this case the source wavefield is going in the upward direction and the receiver wavefield is going in the downward direction, which is as if we change the order of cross-correlation in equation 5. This is why these events map in the time-lag gathers.
with a slope opposite to the primary above the interface. Because the reflected waves only travel in the upper layer, we observe this image above the reflector. In the $R^{tt}(z,\tau)$ image we see two events that map with similar slope, one has the exact opposite slope as the one shown by the primary reflection, the other has a slightly higher slope (therefore indicating faster velocity) and corresponds to the interaction between head-waves produced by the velocity discontinuity and the reflected wavefields. In the time-lag gathers the slope of the primaries is very different from the backscattered events slope. Kaelin and Carvajal (2011) use the slope difference to filter the backscattered events in this domain and to extract the conventional image from the filtered extended image $R(x)=R(x,\tau=0)$.

### 2.2.2 Space-lag common image gathers

Figure 1(b) shows a space-lag gather for the various combinations of the source and receiver wavefield components. We note that with the correct velocity model, both primaries and backscattered events map to $\lambda_x=0$ since the velocity used for imaging is correct. Figure 5(a) shows the $R^{rr}(z,\lambda_x)$ image with the energy correctly focused at $\lambda_x=0$. Figures 5(b) and 5(c) show the backscattered events $R^{tr}(z,\lambda_x)$ and $R^{rt}(z,\lambda_x)$ in the space-lag gathers, which also map toward $\lambda_x=0$. Figure 5(d) shows the image coming from the reflected wavefields $R^{rr}(z,\lambda_x)$; in this case the events are visible only above the reflector because the waves travel only in the first layer.

### 2.2.3 Common-image point gathers

The events involving backscattered energy are also visible in CIP gathers. Figure 1(c) shows a CIP extracted at $\varepsilon=(5,1.5)\text{km}$. Figure 6(a) shows the CIP for the transmitted wavefields $R^{tt}(x,\lambda,\tau)$. The energy focuses at zero lag for the $\tau-\lambda_x$ panel. The $\lambda_x-\tau$ shows a kink produced by the abrupt change in velocity for the primaries, which are mapped at negative $\tau$. Figure 6(c) shows the $R^{rr}(x,\lambda,\tau)$ image, and we can see a change in the $\lambda_x-\tau$ plane, where the backscattered energy is mapped to positive lags. Figure 6(b) shows the complementary backscattered energy that is mapped to negative $\lambda_x$ and positive $\tau$ lags. The CIP from the reflected wavefields, Figure 6(d), shows weak energy concentrated at zero lags.

### 3 SENSITIVITY TO VELOCITY ERRORS

In this section, we test the dependency of the backscattered energy on velocity errors using extended images. In the previous sections we have explained the concept of wavefield synchronization for correct velocity, which implies that for correct velocity the backscattered energy maps toward zero lags. Here, we analyze the behavior of backscattered events in the presence of velocity errors. We test the sensitivity of the backscattered events with the same synthetic data discussed previously. In this case, we construct the images with different models characterized by a constant error varying from -12% to +12% in layer 1. We keep the interface consistent with the velocity used for imaging, i.e. we assume that the interface producing backscattered energy is placed in the model according to the velocity in layer 1. Figures 7(a) to 7(i) show time-lag gathers as a function of the velocity error. The backscattered energy is still mapped vertically, but away from $\tau=0$. The backscattered events in the time-lag gathers show a kinematic error, i.e. these events move from positive $\tau$ for negative errors to negative $\tau$ values for positive errors. Figures 8(b) to 8(i) show a similar display for space-lag gathers. In this case, both, backscattered and primary energy map away from $\lambda_x=0$ when we introduce an error in the model. In space-lag gathers, the backscattered energy maps symmetrically away from zero lag with incorrect velocities. Finally, Figures 9(b) to 9(i) show the sensitivity of CIPs to velocity errors. For incorrect velocity, the events move away from zero-lags. In the CIPs, even when constructed with incorrect velocity, the primary reflections go trough zero-lag, and the events in the $\tau-\lambda_x$ plane show moveout (i.e. the energy not mapped symmetrically respect zero lag). The velocity errors split the backscattered energy in the $\lambda_x-\tau$ plane, some
Understanding RTM backscattering

Figure 5. Illustration of the linearity of the space-lag extended imaging condition. We can divide Figure 8(e) in four images, $R_{tt}'(z, \lambda_x)$ (a), $R_{rt}'(z, \lambda_x)$ (b), $R_{tr}'(z, \lambda_x)$ (c) and $R_{rr}'(z, \lambda_x)$ (d), corresponding to the correlation of the transmitted and/or reflected components of the source and receiver wavefields.

Figure 6. Illustration of the linearity of the extended imaging condition for a common image point. We can decompose a CIP, Figure 9(e), in four images $R_{tt}'(\lambda, \tau)$ (a), $R_{rt}'(\lambda, \tau)$ (b), $R_{tr}'(\lambda, \tau)$ (c), $R_{rr}'(\lambda, \tau)$ (d) corresponding to the correlation between transmitted and/or reflected components of the source and receiver wavefields.

The energy goes through zero space-lag while other part of the energy does not.

We could use the information contained in the extended images to design objective functions (OF) that exploit the presence of backscattered events. Minimizing such OF, e.g. by wavefield tomography, optimizes the sharp interface positioning (e.g. the top of salt) and the sediments velocity above it. A straightforward approach based on differential semblance optimization (Shen et al., 2003) can be adapted to use the backscattered energy seen away from zero lags by defining the objective functions for time-lag gathers

$$J_{\tau} = \frac{1}{2} \| P(\tau) \left[ R_{tr}'(x, \tau) + R_{rt}'(x, \tau) \right] \|_2^2,$$

and for the space-lag gathers,

$$J_{\lambda_x} = \frac{1}{2} \| P(\lambda_x) \left[ R_{tr}'(x, \lambda_x) + R_{rt}'(x, \lambda_x) \right] \|_2^2.$$
Figure 7. Model error sensitivity with time-lag gathers: (a) -12%, (b) -9%, (c) -6%, (d) -3%, (e) 0%, (f) +3%, (g) +6%, (h) +9% and (i) +12% velocity perturbation in the top layer. The maximum energy of the backscattered events occur with correct velocity shown in panel (e).
Figure 8. Model error sensitivity with space-lag gathers: (a) -12%, (b) -9%, (c) -6%, (d) -3%, (e) 0%, (f) +3%, (g) +6%, (h) +9% and (i) +12% velocity perturbation in the top layer. Note that the maximum of backscattered energy happens with the correct velocity shown in panel (e).
Figure 9. Model error sensitivity with CIP gathers: (a) -12%, (b) -9%, (c) -6%, (d) -3%, (e) +0%, (f) +3%, (g) +6%, (h) +9% and (i) +12% velocity perturbation in the top layer.

Here \( P(\tau) = |\tau| \) and \( P(\lambda_x) = |\lambda_x| \) are functions that penalize the backscattered energy away from zero lags, thus defining the residual that we need to minimize trough inversion.

For common image points we can use the objective function

\[
J_c = \frac{1}{2} \| P(\lambda, \tau) [R_{rt}(\lambda, \tau) + R_{rt}(\lambda, \tau)] \|_2^2.
\]  

(9)

Here, \( P(\lambda, \tau) \) is the penalty function for CIPs.

The penalty function is designed to measure the deviation or error between actual extended images and our notion for correct extended images. For CIGs we have a definite criterion, we know that the backscattered energy has to map to zero lag, that is why we can use the absolute value as penalty function. However, for CIPs the penalty operator is more complex. We use the correct CIP as reference for constructing the penalty function \( P(\lambda, \tau) \) similar to the one proposed by (Yang et al., 2012). The correct CIP, shown in Figure 9(e) has the right focusing within the acquisition limitations. More generally, we could use a demigration/migration process to assess correct focusing at a given CIP position, and to infer the shape of the penalty operator.

Figures 10(a) to 10(c) show the penalty functions for
time-lag, space-lag and CIP gathers respectively. The objective functions for our synthetic example are shown in Figures 11(a) to 11(c) for time-lag gathers, space-lag gathers and common image point gathers respectively. One can see that in all three cases the OF minimizes at the correct model. If we want to optimize the model such as we maximize the backscattered events we need to consider two variables: the velocity model and the interface geometry. In our example the sharp interface depends linearly with the velocity model.

4 EXAMPLES

In this section we illustrate the backscattered events visible on extended images constructed based on a modified Sigsbee 2A model (Paffenholz et al., 2002). We modify the model by salt flooding (extending the salt to the bottom of the model) to avoid backscattering from the base of salt, therefore we focus on the reflections from the top of salt only. For this example we fix the receiver array on the surface, and we use 100 shots evenly sampled on the surface to build the image. For the migration model we use the stratigraphic velocity which shows sharp interfaces in the sediment section, in addition the interface corresponding with the top of salt. Figure 12(d) shows the...
conventional image for our modified Sigsbee model, note the strong backscattered energy above the salt.

Figure 12(a) shows a time-lag gather calculated at \( x=19.05 \text{ km} \). We can see that the gather is very complex, but we can easily identify the backscattered energy indicated with letter “a” in Figure 12(a). In this case, the backscattered energy maps directly to \( \tau = 0 \) because we use the correct velocity model. We can also identify the events corresponding to the cross-correlation between reflected waves from the source and receiver side \( R^{rr}(z,\tau) \), indicated with letter “b”. The \( R^{rr}(z,\tau) \) events have positive slope (given by the sediment velocity at the interface) and are visible for \( \tau > 0 \). We can also observe a abrupt change in the slope of the primary reflection corresponding the sediment-salt interfaces at the top of salt indicated with letter “c”.

Figure 12(b) shows a space-lag common image gather extracted at the same location. The backscattered energy maps toward \( \lambda_x = 0 \), indicated with letter “a”. We see again the \( R^{rr}(z,\lambda_x) \) case, indicated with letter “b”, where the energy is mapped away from zero lag. Even though we are using the correct model, we still see energy away from \( \lambda_x = 0 \). This indicates that additional processing is needed before we can use space-lag gathers for model update with wave equation tomography.

Figure 12(c) shows a common image point extracted at the top of salt interface at \((x,z) = (19.05, 3.4) \text{ km} \). Despite the complexity of this image, we can still identify similar patterns as shown in Figure 9(e). The backscattered events are mapped to \( \tau > 0 \) in the \( \tau - \lambda_x \) plane, indicated with letter “a”. In this plane we can separate with the individual contributions from \( R^{tr}(x,\lambda_z,\tau) \) (which maps to \( \lambda_z < 0 \) and \( \tau > 0 \)), and \( R^{rt}(x,\lambda_z,\tau) \) (which maps to \( \lambda_z > 0 \) and \( \tau > 0 \)), because they are imaged into two different events, whereas in the common image gathers discussed before we cannot differentiate the individual contributions, because both cases map to zero lag. The image of the reflector maps as a point to zero lag in the \( \tau - \lambda_z \) plane (indicated with letter “c”).

Understanding the backscattered energy in the extended images for complex scenarios is the first step in using these events for migration velocity analysis. In this report, we used wavefield decomposition to analyze the patterns of the backscattered energy in conventional and extended images. Although effective, wavefield decomposition can be very costly, specially for 3D models. In practice we need to use filtering of the extended images to isolate the events corresponding to the backscattered energy. How we can effectively implement this operation remains subject to future research.
5 CONCLUSIONS

RTM backscattered events maps to zero lag in the extended images when the velocity is correct. This means that the reflected receiver wavefield travels in perfect synchronization with the source wavefield and vice versa. We demonstrate that the RTM backscattered energy is sensitive to kinematics errors in the velocity model. The backscattered energy in the final image should not be considered as an artifact or a drawback of the imaging method; rather, the backscattered energy should be maximized in the image in order to ensure an optimum velocity model. The analysis of backscattered energy on extended images provide a definite criterion to update velocity models with salt interfaces (i.e. in the Gulf of Mexico). Further tests are needed to develop automated velocity model update based on the maximization of the RTM backscattered energy. Our analysis also shows that some energy is mapped away from zero space-lag when we use RTM. This observation suggest that backscattered events should be taken into account in the design of objective functions in wavefield tomography using primary reflections information.

6 ACKNOWLEDGMENTS

Esteban Díaz would like to thank to Mariana Carvalho and Tongning Yang for insightful discussions with the common image point gathers analysis. The reproducible numeric examples in this paper use the Madagascar open-source software package freely available from http://www.reproducibility.org. This research was supported in part by the Golden Energy Computing Organization at the Colorado School of Mines using resources acquired with financial assistance from the National Science Foundation and the National Renewable Energy Laboratory.

REFERENCES

Optimization of time reversal focusing through deconvolution

T.J. Ulrich\textsuperscript{1}, Johannes Douma\textsuperscript{2}, Brian E. Anderson\textsuperscript{1}, and Roel Snieder\textsuperscript{2}

\textsuperscript{1} Geophysics Group EES-17, MS D443, Los Alamos National Laboratory, Los Alamos, New Mexico 87545
\textsuperscript{2} Center for Wave Phenomena, Colorado School of Mines, Golden, Colorado 80401

ABSTRACT

In this study a technique is demonstrated to optimize the ability of time reversal to both spatially and temporally focus, or compress, elastic wave energy. This method allows for optimization of time reversal focusing for any number of channels by calculating the required time reversed signal to be rebroadcast from the time reversal mirror element using only the original recorded scattered signal from that element. The speed, simplicity and robust nature of this process for any number of time reversal mirror elements, including only a single channel, provide the ability to enhance applications currently using time reversal for tasks such as communications, nondestructive evaluation, lithotripsy and imaging, among others. It is also demonstrated theoretically and numerically that good temporal focusing implies that the radius in the spherically symmetric part of the spatial focus is small.

Key words: time reversal, signal processing, communications

1 INTRODUCTION

Time Reversal (TR) in acoustics has been the focus of much research (Parvulescu, 1961; Fink, 1997; Larmat et al., 2010). This has led to many applications in a wide variety of fields, such as medicine, communications, nondestructive evaluation (NDE), seismology, etc. In many of these applications, the ability to use TR depends upon its ability to precisely compress the measured scattered waveforms to a point in both time and space, ideally a delta function, \( \delta(t) \), which is commonly referred to as TR focusing. The desire to enhance the TR process to focus wave energy has lead researchers to develop a variety of ways to accomplish this. Some techniques use arrays of input transducers, measure the wave field with an array near the desired focal spot, and then optimize the spatial and temporal focusing (Tanter et al., 2000, 2001; Montaldo et al., 2004; Vignon et al., 2006; Gallot et al., 2011; Bertaix et al., 2004; Roux and Fink, 2000; Aubry et al., 2001; Jonsson et al., 2004). Other methods use an array of input transducers and optimize the temporal focusing at an output transducer (Daniels and Heath, 2005; Qiu et al., 2006; Blomgren et al., 2008; Zhou et al., 2006; Zhou and Qiu, 2006). Some of these techniques require, or at least benefit greatly from, large arrays, while others may enhance temporal focusing at the expense of spatial localization.

We explore a simple approach and use deconvolution, a primitive, though robust, version of the inverse filter (Tanter et al., 2000; Gallot et al., 2011), to give an input signal that gives a delta function \( \delta(t) \) as an output at the focal point, and show laboratory experiments with ultrasound that the temporal focusing thus obtained is superior to that obtained by using time reversal. By scanning the sample near the focal point, we find that this improved temporal focusing is accompanied by a better spatial focusing. We then show theoretically that improved temporal focusing implies enhanced spatial focusing for the spherical average of the focus. The method described in this paper enhances both spatial and temporal focusing for any number of array elements, including a single channel. In contrast to previous applications of the inverse filter (Tanter et al., 2000, 2001; Gallot et al., 2011), we also explore the effectiveness of using deconvolution for small time reversal mirrors (TRM), i.e., using 1-8 channels, and in the case of limited available bandwidth.

2 DECONVOLUTION AS AN INVERSE FILTER

The time reversal process relies on the assumption that from a recorded impulse response, or Green function
Going back to eq. (1), we can rewrite this (using a convolution notation, rather than the integral form) as

$$f(t) = g(t) * r(t) \approx \delta(t) \ ,$$  \hspace{1cm} (2)

where $f(t)$ is the acoustic wave signal measured at the source, $r(t)$ is the signal sent from the source, and $g(t)$ is the signal that is back propagated to the source. Here we restrict ourselves to looking only at signals between the two points $A$ and $B$ and thus have removed them from the notation. We also change from a Green function notation so as not to imply that we necessarily have an infinite bandwidth available. From eq. (2), we aim to approximate the focused signal $f(t)$ to a delta function by assuming that the recorded impulse response $r(t)$ contains information enough to produce an optimized signal $g(t)$ to rebroadcast in our TR focusing procedure.

Deconvolution equates to inverse filtering by transforming to the frequency domain, thus eq. (2) becomes

$$f(\omega) = g(\omega) \times r(\omega) \approx 1 \ ,$$  \hspace{1cm} (3)

which can easily be manipulated to provide $g(\omega)$ simply from measuring $r(\omega)$ (or rather from the FFT of $r(t)$). As such, we can now calculate the optimized signal which should produce the most impulse-like TR focus from

$$g(\omega) = \frac{1}{r(\omega)} = \frac{r^*(\omega)}{|r(\omega)|^2}.$$  \hspace{1cm} (4)

This expression gives a mathematical expression for $g(\omega)$, but this may not be practical for experimental use in the event that there is a limited bandwidth, or more precisely if $r = 0$ at any frequency. To avoid this

we simply add a constant to the denominator to ensure that we never divide by 0, hence we replace eq. 4 by

$$g(\omega) = \frac{r^*(\omega)}{|r(\omega)|^2 + \epsilon} \ .$$  \hspace{1cm} (5)

where $\epsilon$ is a constant related to the original received signal as

$$\epsilon = \gamma \times \text{mean}(|r(\omega)|^2) .$$  \hspace{1cm} (6)

We used an arbitrary value of $\gamma = 1$ for the experiments described in the next section.

As the above derivation has been performed in the frequency domain, it is necessary to transform back to the time domain to recover the optimized signal, $g(t)$ to be used in the time reversal experiments. It is worth noting that this procedure produces a $g(t)$ that can be used directly for rebroadcasting, i.e., there is no need to perform a TR operation $g(t) \rightarrow g(-t)$ to obtain a focused signal. While this completes the deconvolution optimization procedure used in this paper, it is only a portion of the inverse filter procedure as defined by Tanter et al. (2000), and utilized most recently by Gallot et al. (2011). Performing the full inverse filter procedure requires the singular value decomposition (SVD) of the matrix of filtered waveforms. For reasons that will be addressed in the following section this step is neither necessary nor appropriate for this study.

\section{Experimental Validation}

The purpose of this study is to optimize TR in typical experimental situations, specifically those where only a few channels are used with reverberant signals in a closed cavity. With this in mind, it is prudent to test the methodology experimentally, as presented above. To do this, we used a sample made of aluminum, approximately 1000cm$^3$ volume though irregularly shaped. A
source transducer (2 cm PZT disk) was attached to the sample using epoxy. This defines, and fixes, our source position. The receiver, a non-contact laser vibrometer (PSV 301, OFV 5000 controller, Polytec Inc.) was positioned at a point B on the surface. As we cannot perform a classical TR experiment with a laser vibrometer, since it cannot act as both receiver and source, we performed reciprocal TR experiments, i.e., where the original source and back propagation signals are emitted from the same location/transducer, thus focusing the wave energy to the original receiver location. This use of TR in a reciprocal sense is now a widely used form of TR, and indeed is the method originally used by Parvelescu (Parvulescu, 1961) in the first known demonstration of TR in acoustics (also known as phase conjugation). It is due to the use of reciprocal TR that SVD is unnecessary. The SVD step of the inverse filter procedure is applicable to the classical TR procedure as its purpose is to remove correlated noise that may exist simultaneously in all of the received signals. Reciprocal TR requires that each received signal \( r(t) \) is obtained individually, i.e., not simultaneously. For this reason there is no expected correlation of noise from one measurement of \( r(t) \) to the next.

The original impulse, a 5\( \mu \)s shaped toneburst of 200kHz, was broadcast from the source transducer using a 12-bit arbitrary waveform generator with a conversion rate of 10MHz. The response to this impulse was recorded by the laser vibrometer (5mm/s/V, 250kHz bandwidth) using a 14-bit digitizer also sampling at 10MHz. The generator and digitizer were synched such that the source impulse was centered in a 3.2768ms window, thus the first portion of the received signal \( r(t) \) contains no causal signal. This also means that the TR focus will occur at 1.6384ms during the rebroadcast.

In fig. 1, both the reversed original response \( r(-t) \) and the result of the optimization procedure \( g(t) \) can be seen. Each of these signals were used independently to achieve a TR focus. These focused signals, corresponding to the back propagation of these signals, can be found in fig. 2 (a). There are a few key points to be noted in these two figures. First, note the acausal signal, i.e., signal appearing after \( t = 1.64 \text{ms} \), in the optimized signal. This portion of the signal is necessary for the reconstruction of a symmetric focal signal. Arbitrarily setting \( g(t > 1.64 \text{ms}) = 0 \) does not affect the impulse-like focus at 1.6384ms in fig. 2 (a). However, it does destroy the symmetry that is seen in fig. 2 (a) for the optimized focus.

In fig. 2, the two important features of note are: i) the temporal compression, and ii) maximum achievable amplitudes. The normalization of the focal signals (and the original source signal) shown in fig. 2 (a) provide a clear demonstration of the enhanced temporal compression. This can be quantified by calculating the total energy near the focal time (50 samples centered at 1.6384ms, corresponding to the duration of the original source) and normalizing by the total energy in the signal. Performing this metric for these signals, we find that the original source has 100% of the energy in this 50 sample window, while the standard and optimized focused signals contain \( \sim 15\% \) and 70% of the total energy respectively. Fig. 2 (b) shows further evidence in the lack of energy in the optimized signal (red) at all times away from the focal window, whereas the standard TR focal signal exhibits temporal side lobes symmetrically around the focal time. The second point to be noted from these figures can be seen in fig. 2 (b). The maximum achieved focal amplitude for the optimized process is approximately one half of that achieved by the standard TR process, a fact that we observe in all tests of this technique for all numbers of channels 1-8.

While figs. 1 and 2 effectively illustrate the optimization of temporal compression, albeit at the expense of the amplitude, they do not in any way address the spatial extent of the focus. To explore this, the same laser vibrometer was used in a scanning mode to measure the wavefield in a \( 5 \times 5 \text{cm}^2 \) region around the focus point (i.e., location \( B \)) with a resolution of 2mm. The images in fig. 3 show the wavefield measured at the time of focus (\( t = 1.6384 \text{ms} \)) for both the standard and optimized TR process. The standard TR process in fig. 3 (a) shows the typical focal spot of approximately \( \lambda/2 \) with a fringe, or ring, appearing at approximately \( 2\lambda \) from the focal point. Using the optimized signal \( g(t) \) from fig. 1, on the other hand, appears to have a smaller spatial extent, see fig. 3 (b), but most notably lacks any structure away from the central focus. Normalizing the two images and performing a simple subtraction (standard - optimized) verifies the more peaked nature of the optimized...
enhances the energy compression over the standard TR procedure. When performing these same tests for variable number of channels (1-8) the results are consistent showing the maximal compression is achieved at approximately 4-5 TRM elements (for either deconvolution or standard TR), after which time total amplitude increases linearly with number of channels (not shown) but further enhancement to focusing is minimal (fig. 4).

To accomplish this study as a function of TRM elements, the same experimental procedure as previously described was used for up to 8 individual transducers affixed at 8 unique locations on the aluminum sample. For 1 channel, the values in Fig. 4 result from an average of the temporal and spatial metrics across the results from the 8 individual channels. The results for N channels come from summing the data of N channels to produce the two metrics and then averaging these metric values across various combinations of N of the 8 channels. The results for 8 channels is a summation of all 8 individual data to produce one value each for the temporal and spatial metrics.

In conclusion, the main objective for the use of de-
convolution was to improve the temporal focusing. We showed experimentally that simple deconvolution can achieve this, albeit at the expense of the amplitude. We then discovered and showed experimentally that, not only did the method improve the temporal focus, the spatial focus was also improved. In the next section, we present one explanation for both improved spatial and temporal focusing.

4 IMPROVED TEMPORAL FOCUSING LEADS TO IMPROVED SPATIAL FOCUSING

In this section, we will show that better temporal focusing implies better spatial focusing for the spherical average of the focus. We begin by first considering the wave field near its focal spot at \( r = 0 \) and consider the medium to be locally homogeneous in that region. The solution of the Helmholtz equation in a homogeneous medium can be written as

\[
p(r, \theta, \varphi, \omega) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} a_{lm} j_l (kr) Y_{lm}(\theta, \varphi),
\]

see table 8.2 of ref. Arfken and Weber (2001). In this expression \( j_l \) denotes the spherical Bessel function, \( Y_{lm} \), the spherical harmonics, and \( k = \omega/c \). According to expressions (11.147) and (11.148) of ref. Arfken and Weber (2001), \( j_l(0) = 0 \) for \( l \geq 1 \) and \( j_0(0) = 1 \). This means that at the focal point \( r = 0 \) only the terms \( l = m = 0 \) contribute. Using that \( Y_{00} = 1/\sqrt{4\pi} \) (table 12.3 of ref. Arfken and Weber (2001)), this means that at the focal point

\[
p(r = 0, \theta, \varphi, \omega) = \frac{a_{00}}{\sqrt{4\pi}}.
\]

The properties of the wave field at the focal point thus only depend on the coefficient \( a_{00} \). Since the \( l = m = 0 \) component of the spherical harmonics expansion gives the spherically symmetric component of the wave field, the properties of the wave field at the focal point can only bear a relation to the spherically symmetric component of the wave field. The properties of temporal focusing can thus only be related to the spherically symmetric component of the spatial focusing.

Because of this, we restrict ourselves to the spherically symmetric component \((l = m = 0)\) of the wave field at the focal point. Using that \( j_0(kr) = \sin(kr)/kr \), the spherically symmetric component of expression (7) is given by

\[
p(r, \omega) = p_0 e^{-ikr} - e^{ikr},
\]

with \( p_0 = -a_{00}/(2ik\sqrt{4\pi}) \). The coefficient \( p_0 \) depends on frequency. Using the Fourier convention \( f(t) = \int p_0(\omega)e^{-i\omega t}d\omega \), and using that \( k = \omega/c \), expression (9) corresponds, in the time domain, to

\[
p(r, t) = \frac{f(t + r/c) - f(r - r/c)}{r}.
\]

In this expression, \( f(t + r/c) \) denotes the wave that is incident on the focus, and \( f(t - r/c) \) the outgoing wave once it has passed through the focus. The field at the focus follows by Taylor expanding \( f(t \pm r/c) \) in \( r/c \) and taking the limit \( r \to 0 \), this gives

\[
p(r = 0, t) = \frac{2}{c} f'(t).
\]

In this expression and the following, the prime denotes the time derivative. Expression (11) states that the wave field at the focus is the time derivative of the incoming wave field.

Equation (11) gives the temporal properties of the focus. In order to get the spatial properties, we consider the wave field near the focal point at time \( t = 0 \). Setting \( t = 0 \) in expression (10), and making a third order Taylor expansion of \( f(\pm r/c) \) gives

\[
p(r, t = 0) = \frac{2}{c} f'(0) + \frac{r^2}{3c^3} f''''(0).
\]

Using equation (11) to eliminate \( f \), gives

\[
p(r, t = 0) = p(r = 0, t = 0) + \frac{1}{6c^3} p'''(r = 0, t = 0) r^2.
\]

This is a parabolic approximation for the wave field near the focus, with coefficients related to the wave field and the focus and its second time derivative. The radius \( R \) of the focal spot can be estimated by setting the left hand side of equation (13) equal to zero, this gives

\[
R = \sqrt{6c^3} \int \frac{-p(r = 0, t = 0)}{p''(r = 0, t = 0)}
\]

A good temporal focusing means that the temporal curvature of the wave field at the focal point is strong, this means that \(-p''/p\) is large, and that the radius \( R \) is small. Good temporal focusing thus implies that the radius in the spherically symmetric part of the spatial focus is small.

5 CONCLUSION

Here we have introduced a simple method for determining the optimal signal to be used in a TR experiment in order to maximally compress the focus in both time and space. This method has been experimentally verified, albeit without fully varying all possible parameters. A more detailed study is currently underway to quantify the effects due to variations in available bandwidth,
and material properties (e.g., homogeneous vs. heterogeneous media, range of attenuations, wave speeds, etc.)

While temporal and spatial focusing have been enhanced with this method, there is a cost in the maximum achieved amplitude. This fact, coupled with the effect of the acausal portion of $g(t)$ may lead one to think simplistically about $g(t)$ as containing a signal to focus the energy at the appropriate time and place, and another signal that, simultaneously, actively cancels noise at all other times and locations. Thus the optimized focus is cleaner and more impulse-like, but not as strong, as some of the energy being transmitted is being exerted in suppressing sidelobes, spatially and temporally.

We have also shown, theoretically and numerically, that if one has good temporal focusing, the radius in the spherically symmetric part of the spatial focus is small. This concept can turn out to be useful in future research because it allows one to work on improving the temporal focusing in order to improve the spatial focusing.

ACKNOWLEDGMENTS

Work supported by Los Alamos National Laboratory Institutional Support (LDRD).

REFERENCES


Multiparameter TTI tomography of P-wave reflection and VSP data

Xiaoxiang Wang & Ilya Tsvankin
Center for Wave Phenomena, Geophysics Department, Colorado School of Mines, Golden, Colorado 80401

ABSTRACT
Transversely isotropic models with a tilted symmetry axis (TTI media) are widely used in depth imaging for complex geologic structures. Here, we present a modification of a previously developed 2D P-wave tomographic algorithm for building heterogeneous TTI models and apply it to synthetic data. The symmetry-direction velocity $V_{P0}$, anisotropy parameters $\epsilon$ and $\delta$, and the symmetry-axis tilt $\nu$ are defined on a rectangular grid. To ensure stable reconstruction of the TTI parameter fields, reflection data are combined with walkaway VSP (vertical seismic profiling) traveltimes in joint tomographic inversion. To improve the convergence of the inversion algorithm, we propose a three-stage model-updating procedure that gradually relaxes the constraints on the spatial variations of the anisotropy parameters, while the symmetry axis is kept orthogonal to the reflectors. Only at the final stage of the inversion the parameters $V_{P0}$, $\epsilon$, and $\delta$ are updated on the same grid. We also incorporate geologic constraints into tomography by designing regularization terms that penalize parameter variations in the direction parallel to the interfaces. First, the reflection tomography without borehole constraints is tested on a model that includes a bending TTI layer with a wide range of dips. Then we examine the performance of the regularized joint tomography of reflection and VSP data for two sections of the BP TTI model that contain an anticline and a salt dome. The TTI parameters in the shallow part of both sections (down to 5 km) are well-resolved by the three-stage model-updating process. Due to the limited constraints from reflection events and sparse coverage of VSP rays at depth, the velocity field in the deeper part of the section is estimated with larger errors. These results provide useful guidance for building accurate TTI models for prestack depth imaging.

1 INTRODUCTION
Prestack depth imaging for complex geologic environments (including fold-and-thrust belts and subsalt plays) requires anisotropic velocity models such as transverse isotropy with a vertical (VTI) or tilted (TTI) axis of symmetry. Therefore, conventional isotropic reflection tomography for isotropic media (Stork, 1992) need to be extended to heterogeneous anisotropic media (Campbell et al., 2006; Woodward et al., 2008). Because P-wave reflection traveltimes do not provide sufficient constraints for resolving all relevant TTI parameters, it is necessary to use additional information (e.g., borehole data) to reduce the nonuniqueness of the inverse problem (Moric et al., 2004; Tsvankin, 2005; Bakulin et al., 2010a).

P-wave velocity in TTI media is controlled by the velocity $V_{P0}$ in the symmetry-axis direction, anisotropy parameters $\epsilon$ and $\delta$, and the orientation of the symmetry axis (in 2D defined by the tilt $\nu$ from the vertical). Zhou et al. (2011) develop multiparameter reflection tomography for TTI media and apply it to synthetic and field data. They find that simultaneous estimation of all three relevant parameters ($V_{P0}$, $\epsilon$, and $\delta$) helps flatten the common-image gathers (CIGs) better than single-parameter (only $V_{P0}$) inversion. Zhou et al. (2011) also confirm that trade-offs between the TTI parameters cannot be eliminated using only P-wave reflections, and point out the importance of additional constraints from well data. However, they do not carry out joint inversion of reflection and borehole data by including, for example, VSP (vertical seismic profiling) traveltimes.

Nonuniqueness of the inversion of reflection data can be mitigated by regularization which imposes a priori constraints on the estimated model (Engl et al., 1996). Wang and Tsvankin (2011; hereafter, referred to as Paper I) employ Tikhonov (1963) regularization to smooth the velocity field both horizontally and vertically. Fomel (2007) develops so-called “shaping regular-
ization” designed to steer the velocity variations along geologic structures (e.g., layers). His mapping (shaping) operator is integrated into the conjugate-gradient iterative solver. Using the steering-filter preconditioner (Clapp et al., 2004) similar to shaping regularization, Bakulin et al. (2010b) perform joint tomographic inversion of P-wave reflection data (from horizontal and dipping reflectors) and check-shot traveltimes for VTI media. They conclude that in the vicinity of the well it is possible to resolve the vertical variation of all three relevant parameters ($V_{p0}$, $\epsilon$, and $\delta$).

In Paper I, we develop a 2D ray-based tomographic algorithm for iteratively updating the parameters $V_{p0}$, $\epsilon$, and $\delta$ of TTI media defined on rectangular grids (the symmetry axis is set orthogonal to the imaged reflectors). Synthetic tests for a simple model with a “quasifactorized” TTI syncline (i.e., $\epsilon$ and $\delta$ are constant inside the TTI layer, while the tilt $\nu$ may vary spatially) demonstrate that stable parameter estimation requires strong smoothness constraints or additional information from walkaway VSP traveltimes.

Here, we first introduce the objective function that contains the residual moveout in CIGs and the VSP traveltimes computed for each trial model and included in the objective function:

$$F(\Delta \lambda) = \|A \Delta \lambda - b\|^2 + c_{\text{vsp}}^2 \|E \Delta \lambda - d\|^2 + R(\Delta \lambda),$$  \hspace{1cm} (1)

where the vector $\Delta \lambda$ represents the parameter updates, the elements of the matrix $A$ are the traveltime derivatives with respect to the medium parameters at each grid point ($A$ is computed analytically along the ray-paths), $b$ is a vector containing the residual moveout in CIGs, the matrix $E$ is composed of VSP traveltime derivatives, and the vector $d$ is the difference between the observed and calculated VSP traveltimes for each source-receiver pair. The regularization term $R$ in equation 1 has the form:

$$R(\Delta \lambda) = \zeta^2 \|\Delta \lambda\|^2 + \zeta_1^2 \|L_1(\Delta \lambda + \lambda^0)\|^2 + \zeta_2^2 \|L_2(\Delta \lambda + \lambda^0)\|^2,$$  \hspace{1cm} (2)

where $\zeta^2 \|\Delta \lambda\|^2$ restricts the magnitudes of parameter updates that have small derivatives in the matrices $A$ and $E$, the operators $L_1$ and $L_2$ are designed to make parameter variations more pronounced in the direction normal to the interfaces, and $\zeta_1$, $\zeta_2$, and $\zeta_3$ are the regularization coefficients/weights. In 2D, the normal direction of a reflector is defined by the dip angle with the vertical, which is computed from the depth image using Madagascar program “sfdip.” Then the symmetry-axis tilt $\nu$ at each grid point is set to be equal to the corresponding dip.

To construct the matrix $L_1$, we first compute two components of the gradient vector $\nabla \lambda$ from the following finite-difference approximation:

$$\lambda'(x) = \frac{\lambda(\text{\textbf{x}} + dx) - \lambda(\text{\textbf{x}} - dx)}{2dx} + O[(dx)^2],$$  \hspace{1cm} (3)

$$\lambda'(z) = \frac{\lambda(\text{\textbf{z}} + dz) - \lambda(\text{\textbf{z}} - dz)}{2dz} + O[(dz)^2],$$  \hspace{1cm} (4)

where $\lambda$ is the parameter ($V_{p0}$, $\epsilon$, or $\delta$) at the grid point with the coordinates $x$ and $z$, and $dx$ and $dz$ are the cell dimensions. Since the dip field yields the vector $\text{n}$ orthogonal to reflectors, we minimize the norm of the cross-product $\|\text{n} \times \nabla \lambda\|$ at all grid points, which is equivalent to aligning the direction of the largest parameter variation with $\text{n}$ and restricting the variations along interfaces. To include more cells, we can use a higher-order finite-difference approximation:

$$\lambda'(x) = \frac{[-\lambda(x + 2dx) + 8\lambda(x + dx) - 8\lambda(x - dx) + \lambda(x - 2dx)]/(12dx) + O[(dx)^4]}{12dx},$$  \hspace{1cm} (5)

$$\lambda'(z) = \frac{[-\lambda(z + 2dz) + 8\lambda(z + dz) - 8\lambda(z - dz) + \lambda(z - 2dz)]/(12dz) + O[(dz)^4]}{12dz}.$$  \hspace{1cm} (6)

Similarly, the finite-difference approximation of the Laplacian operator $\nabla^2 \lambda$ has two components:

$$\lambda''(x) = \frac{\lambda(x + dx) - 2\lambda(x) + \lambda(x - dx)}{(dx)^2},$$  \hspace{1cm} (7)

$$\lambda''(z) = \frac{\lambda(z + dz) - 2\lambda(z) + \lambda(z - dz)}{(dz)^2}.$$  \hspace{1cm} (8)

Then we compute the cross-product $\text{n} \times \nabla^2 \lambda$ and minimize its norm to mitigate the parameter variations in the direction parallel to the interfaces.

Defining the TTI parameters on a relatively small grid results in a large number of unknowns, and the Fréchet matrices $A$ and $E$ in equation 1 are sparse (i.e., only nonzero or large elements are stored due to the limited computer memory). To solve the large sparse linear system of equations in an efficient way, we employ the parallel direct sparse solver (PARDISO) from Intel Math Kernel Library (MKL, http://software.intel.com/en-us/articles/intel-mkl/).
As described in Paper 1, the anisotropic velocity field is iteratively updated starting from an initial model that may be obtained from stacking-velocity tomography at borehole locations (Wang and Tsvankin, 2010). In the first few iterations, the symmetry-direction velocity \( V_{0} \) is typically inaccurate and simultaneous inversion for all TTI parameters may result in unacceptably large updates for the anisotropy parameters \( \epsilon \) and \( \delta \). If the anisotropy parameters are moderate (\( \epsilon < 0.25 \) and \( \delta < 0.15 \) in our tests), it is convenient to fix them temporarily at the initial values (typically small) and limit the updates to the velocity \( V_{0} \). At the second stage of parameter updating, the model is divided into several layers based on the picked reflectors, and the anisotropy parameters are assumed to be constant within each layer. The velocity \( V_{0} \) is then updated on a grid, while \( \epsilon \) and \( \delta \) change in each layer (i.e., the inversion for \( \epsilon \) and \( \delta \) is layer-based). Such a “quasi-factorized” assumption is equivalent to strong smoothing of the anisotropy parameters and may help resolve all TTI parameters if \( V_{0} \) is a linear function of the spatial coordinates and at least two distinct dips are available (Behera and Tsvankin, 2009). At the third and last stage of velocity analysis, the parameters \( \epsilon \) and \( \delta \) are updated on the same grid as that for \( V_{0} \), to allow for more realistic treatment of heterogeneity. Still, because P-wave kinematics are less sensitive to the anisotropy parameters than to the symmetry-direction velocity, the \( \epsilon \)– and \( \delta \)–fields in function 2 should be regularized with larger weights.

3 SYNTETIC EXAMPLES

3.1 TTI thrust sheet

First, the tomographic algorithm is tested on the synthetic data of Zhu et al. (2007), whose model (simulating typical structures in the Canadian Foothills) includes a TTI thrust sheet embedded in an otherwise isotropic, homogeneous medium (Figure 1). P-wave reflection data were generated by an anisotropic finite-difference code. We used sources and receivers placed every 60 m with the maximum offset reaching 1980 m. Because the exact model geometry (i.e., the interface positions) is unavailable, we cannot provide comparisons of our migration results with the correct model.

The initial model for reflection tomography includes two horizontal isotropic layers (Figure 2a). Although the P-wave velocity in the isotropic background is set to the correct value, ignoring transverse isotropy in the bending layer causes noticeable residual moveout in common-image gathers (Figure 2b) and a strong distortion of the imaged reflector beneath the thrust sheet (Figure 2c). In the velocity-updating process, the velocity \( V_{0} \) is defined on a square (100 m \( \times \) 100 m) grid.

Because of the relatively simple model geometry, it is not necessary to follow the three-stage inversion procedure introduced above. Based on the picked reflectors, the section is divided into two isotropic blocks and a TTI layer sandwiched between them. Each layer/block is assumed to be “quasi-factorized” TTI with constant anisotropy parameters \( \epsilon \) and \( \delta \) (i.e., we only use the second step of the updating procedure). The symmetry axis is taken perpendicular to the reflectors, with the tilt changing during the updates. Because there is only a single horizontal reflector on the right side of the model (Figure 1), the parameters \( \epsilon \) and \( \delta \) above it cannot be resolved solely from P-wave reflection data. Therefore, both \( \epsilon \) and \( \delta \) in the block to the right of the TTI layer are set to zero.

With a general smoothing operator (Wang and Tsvankin, 2011) instead of \( L_{1} \) and \( L_{2} \) in function 2, the velocity in the TTI layer is partially recovered (Figure 3a) after 12 iterations. Because \( V_{0} \) is updated on a relatively fine grid, flattening the CIGs along three reflectors (and just one reflector for the block on the right side of the model) is insufficient for recovering the velocity field, even when regularization is applied. For example, there is noticeable heterogeneity in each block that does not exist in the correct model.

Nevertheless, the constraints provided by a wide range of dips in the TTI thrust sheet and the correct assumption about the spatial variation of \( \epsilon \) and \( \delta \) helps accurately resolve both anisotropy parameters (Figure 3b and 3c). The error in \( \epsilon \) in the TTI layer (0.03) is somewhat larger than that in \( \delta \) (-0.01) because of the small offset-to-depth ratio, which is close to unity for the bottom reflector. As shown by Behera and Tsvankin (2009), stable estimation of \( \epsilon \) in quasi-factorized TTI media requires long spreads reaching two reflector depths. Despite remaining distortions in \( V_{0} \), the obtained anisotropic velocity field largely removes the residual moveout in the CIGs (Figure 4a) and improves the depth image, especially that of the bottom horizontal reflector (Figure 4b). Additional reflectors or
walkaway VSP data would help refine the velocity field and obtain a more accurate spatial distribution of $V_{P0}$. Another way to suppress spurious spatial variations of $V_{P0}$ is to apply stronger smoothness constraints, which could be based on a priori information about the model.

3.2 BP anticline model

Next, we test the joint tomography of P-wave long-spread reflection data and walkaway VSP traveltimes on a section of the BP TTI model that contains an anticline structure (http://www.freeusp.org/2007_BP_Anisotropy/Benchmark/). The velocity $V_{P0}$ in the actual model is smoothly varying (Figure 7a), except for a small jump at the water bottom. The symmetry axis is set perpendicular to the interfaces (Figure 7b), and the anisotropy parameters $\epsilon$ and $\delta$ change from layer to layer with relatively weak lateral variations compared to those of $V_{P0}$ (Figure 7c and 7d). The depth image produced by Kirchhoff prestack depth migration with the correct velocity model is shown in Figure 8.

Migration velocity analysis (MVA) is applied to CIGs from $x = 40$ km to 61 km with an interval of 150 m (the maximum offset is 10 km). Synthetic VSP data were generated in a vertical “well” placed at location $x_{\text{VSP}} = 51.4$ km with 24 receivers spanning the interval from $z = 4756.25$ m to 5043.75 m every 12.5 m and 24 more receivers evenly placed between 7837.5 m and 8125 m. The VSP sources were located at the...

Figure 2. (a) Initial isotropic model used in velocity analysis for the model from Figure 1. The P-wave velocities in the top and bottom layers are 2740 m/s and 3200 m/s, respectively. (b) The CIGs computed with the initial model and displayed every 0.6 km. (c) The corresponding depth image.

Figure 3. (a) Symmetry-direction velocity $V_{P0}$ estimated at each grid point. The inverted block-based parameters (b) $\epsilon$ and (c) $\delta$. In the TTI thrust sheet, the estimated $\epsilon = 0.19$ and $\delta = 0.07$. In the block to the right of the TTI layer, $\epsilon$ and $\delta$ are set to zero.
2D joint TTI tomography

Figure 4. (a) CIGs after 12 iterations and (b) the corresponding depth image obtained with the parameters from Figure 3. Note that the reflector beneath the thrust sheet has been mostly flattened.

Figure 5. Initial isotropic model with the velocity $V_{P0}$ defined on a 200 m $\times$ 100 m grid. The velocity in the water is set to the correct value.

surface every 50 m from $x = 41.4$ km to 61.4 km (the maximum offset is also 10 km). In addition, check-shot traveltimes were recorded every 50 m from $z = 943.5$ m to 9493.75 m.

To build an initial model, we compute a 1D profile of $V_{P0}$ from the check-shot traveltimes and then obtain the 2D velocity field (Figure 5) by extrapolation that conforms to the picked interfaces. The exact position of the water bottom is assumed to be known, and the velocity of the water layer is fixed at the correct value.

Following the three-stage parameter-estimation procedure described above, first we update only the velocity $V_{P0}$ defined on a rectangular 200 m $\times$ 100 m grid, while keeping the anisotropy parameters $\epsilon$ and $\delta$ set to zero. Then $\epsilon$ and $\delta$ are taken constant in each layer (delineated by the interfaces picked on the image), but updated simultaneously with the velocity. With this “quasi-factorized” TTI assumption, the inverted model (Figure 10) improves the positioning of the reflectors (Figure 12b) and reduces the residual moveout in CIGs (Figure 9a) and the VSP traveltimes misfit.

However, the residual moveout in Figure 12a is not completely removed, mainly because the assumption about the anisotropy parameters does not conform to the actual $\epsilon$– and $\delta$– fields (Figures 7c and 7d). To allow for more realistic spatial variations, at the last stage of parameter updating we estimate the parameters $\epsilon$ and $\delta$ on the same grid as the one used for the velocity $V_{P0}$. However, because the trade-offs between the parameters may cause large errors in $\epsilon$ and $\delta$ defined on small grids, the anisotropy parameters should be more tightly constrained, so the corresponding regularization coefficients should be larger than those for $V_{P0}$.

After five more iterations with all three parameters updated on grids, the velocity $V_{P0}$ above $z = 7$ km is relatively well-recovered with percentage errors in most areas smaller than 4% (Figure 11a). The spatial variations of $\epsilon$ and $\delta$ are partially resolved from the water bottom down to $z = 5$ km (Figure 11c and 11d). The coverage of VSP rays, however, becomes more sparse with depth. Also, the offset-to-depth ratio of P-wave reflections is insufficient to constrain the parameter $\epsilon$ below 7 km, although the maximum offset reaches 10 km. Therefore, the accuracy in $\epsilon$ and $\delta$ decreases in the deep part of the model. The final inverted model (Figure 11) practically removes the residual moveout in CIGs (Figure 13a) (except for the locations close to the left and right edges due to poor ray coverage), and the reflections are better focused (Figure 13b), especially above $z = 7$ km.

An important parameter that influences the accuracy of the reconstructed velocity model is the symmetry-axis tilt $\nu$, which is computed directly from the depth image. Poorly constrained TTI parameters in the deep part of the model yield a strongly distorted image, which produces large errors in the estimated values of $\nu$. The obtained tilt field is used for the next iteration of MVA, which further distorts the other estimated TTI parameters. Therefore, without sufficient constraints from deep reflection events and VSP rays, the trade-offs between the tilt $\nu$ and the other TTI parameters increase the uncertainty in velocity analysis at depth.
3.3 BP salt model

The last test is performed for another section of the BP TTI model that includes a salt dome (Figure 14). The strong reflections from the top of the salt dome and the flanks right beneath it are clearly imaged (Figure 15) by Kirchhoff depth migration, but the deeper segments of the flanks are blurred even when the correct model is used. The image quality can be improved with a wavefield-based imaging algorithm, such as reverse time migration (RTM).

The maximum offset (10 km) and the source and receiver intervals (50 m) are the same as in the previous test. The CIGs used for MVA are computed every 150 m from \( x = 16 \) km to 46 km. The dataset contains a vertical “well” at location \( x_{\text{VSP}} = 29.9 \) km to the left of the salt body (Figure 14). Two sets of 24 receivers (one between \( z = 5275 \) m and 5562.5 m and the other between \( z = 8400 \) m and 8687.5 m) were placed at even intervals in the well to record a walkaway VSP survey. The maximum offset for the VSP data is 10 km as well with a source interval of 50 m. The input data also include the check-shot traveltimes obtained every 50 m from \( z = 1743.75 \) m to 9093.75 m.

During the inversion, the water layer and salt body are kept isotropic with the velocities fixed at the correct values. Also, the actual positions of the top and flanks of the salt dome are assumed to be known, and the update is performed only for the sedimentary formations around the salt body. Since ray tracing becomes unstable in the presence of sharp velocity contrasts, we apply 2D smoothing to the velocity model to find the raypaths crossing the salt and then calculate the traveltimes and their derivatives in the original (unsmoothed) model.

Similar to the previous test, an initial isotropic model (Figure 6) is built from check-shot traveltimes. Because the TTI parameters are different on both sides of the salt body, the residual moveout in the CIGs (Figure 16a) is larger to the right of the salt (i.e., further away from the well). Due to the large velocity errors in the initial model, most reflectors are misplaced (Figure 16b).

After two iterations of velocity \( (V_{P0}) \) updating with fixed \( \epsilon \) and \( \delta \) and two more iterations with the “quasi-factorized” TTI model assumption (see above), the estimated parameter fields (Figure 17) produce relatively flat CIGs (Figure 19a) and an improved image (Figure 19b). For the VSP sources placed to the left of the well, the corresponding rays pass through the relatively simple sedimentary section. The VSP rays originated to the right of the well, however, cross the high-velocity salt body, and even small errors in the position of the salt boundary can cause large perturbations of the ray trajectories. Therefore, we assign smaller weights in the objective function to the VSP traveltimes for the sources to the right of the well. Hence, the anisotropic velocity field on the right side of the salt dome has to be determined mostly from the P-wave reflection data, which leads to larger errors in the TTI parameters.

To further reduce the residual moveout in CIGs and the VSP traveltime misfit, the anisotropy parameters are estimated on the same grid as that for \( V_{P0} \). After three more iterations, the velocity (Figure 18a) above \( z = 7 \) km on the left side of the salt body is relatively well-resolved (errors in most areas do not exceed 3%); however, the errors in \( V_{P0} \) on the right side are higher because of the limited constraints from VSP data, as described above. The spatial variations of \( \epsilon \) and \( \delta \) are partially recovered from the water bottom down to \( z = 5 \) km (Figure 18c and 18d). Because of the limited offset-to-depth ratio and poor coverage of VSP rays (especially for the right part of the model) at depth, the anisotropy parameters for grid points below 5 km could not be updated after the first iteration. Using the final model, the residual moveout in CIGs (Figure 20a) and the VSP traveltime misfit are largely reduced, which produces a more accurate image (Figure 20b).
Figure 7. Section of the BP TTI model that includes an anticline (the grid size is 6.25 m × 6.25 m). The top water layer is isotropic with velocity 1492 m/s. (a) The symmetry-direction velocity $V_{P0}$. The black line marks a vertical “well” at $x = 51.4$ km. (b) The symmetry-axis tilt $\nu$. The anisotropy parameters (c) $\epsilon$ and (d) $\delta$.

Figure 8. Depth image produced by prestack migration with the actual parameters from Figure 7. Imaging was performed using sources and receivers placed every 50 m.
Figure 9. (a) CIGs (displayed every 3 km from 41 km to 62 km) and (b) the migrated section computed with the initial model from Figure 5.
Figure 10. Anticline model from Figure 7 updated using the “quasi-factorized” TTI assumption. (a) The symmetry-direction velocity $V_{P0}$ estimated on a 200 m x 100 m grid. (b) The tilt $\nu$ obtained by setting the symmetry axis perpendicular to the reflectors. The inverted interval parameters (c) $\epsilon$ and (d) $\delta$, which are constant within each layer.

Figure 11. Inverted TTI parameters (a) $V_{P0}$, (c) $\epsilon$, and (d) $\delta$ after the final iteration of MVA. All three parameters are estimated on a 200 m x 100 m grid. (b) The symmetry-axis tilt $\nu$ computed from the depth image obtained before the final iteration.
Figure 12. (a) CIGs and (b) the migrated section obtained with the “quasi-factorized” TTI model from Figure 10.
Figure 13. (a) CIGs and (b) the migrated section computed with the final inverted model in Figure 11.
Figure 14. Section of the BP TTI model with a salt dome (the grid size is 6.25 m x 6.25 m). The top water layer and the salt body are isotropic with the P-wave velocity equal to 1492 m/s and 4350 m/s, respectively. (a) The symmetry-direction velocity $V_{P0}$. The vertical “well” at $x = 29.9$ km is marked by a black line. (b) The tilt of the symmetry axis, which is set orthogonal to the interfaces. The anisotropy parameters (c) $\epsilon$ and (d) $\delta$.

Figure 15. Depth image produced with the actual parameters from Figure 14.
Figure 16. (a) CIGs (displayed every 3.25 km from 18 km to 44 km) and (b) the migrated section computed with the initial model in Figure 6.
Figure 17. Salt model from Figure 14 updated using the “quasi-factorized” TTI assumption. (a) The symmetry-direction velocity $V_{P0}$ estimated on a 200 m $\times$ 100 m grid. (b) The tilt $\nu$ obtained from the image. The inverted interval parameters (c) $\epsilon$ and (d) $\delta$, which are constant within each block.

Figure 18. Inverted TTI parameters (a) $V_{P0}$, (c) $\epsilon$, and (d) $\delta$ after the final iteration of joint tomography. All three parameters are estimated on a 200 m $\times$ 100 m grid. (b) The symmetry-axis tilt $\nu$ computed from the depth image obtained before the final iteration.
Figure 19. (a) CIGs and (b) the migrated section obtained with the “quasi-factorized” TTI model from Figure 17.
Figure 20. (a) CIGs and (b) the migrated section computed with the final inverted model in Figure 18.
4 CONCLUSIONS

Although most migration techniques have been extended to TTI media, accurate reconstruction of the anisotropic velocity field remains a difficult problem. Previously we developed an efficient 2D tomographic algorithm for heterogeneous TTI models, with the parameters \( V_{p0}, \epsilon, \delta, \) and the symmetry-axis tilt \( \nu \) defined on a rectangular (in most cases square) grid. While \( V_{p0}, \epsilon, \) and \( \delta \) are updated iteratively in the migrated domain, the tilt field is computed from the depth image by setting the symmetry axis perpendicular to the reflectors.

To resolve the TTI parameters in the presence of spatial velocity variations, here we combined reflection data with walkaway VSP and check-shot traveltimes. Our tomographic algorithm also incorporates useful geologic constraints by appropriately designed regularization. The regularization terms in the objective function allow for parameter variations across layers, but suppress them in the direction parallel to boundaries. In the iterative inversion, such structure-guided regularization also helps propagate along interfaces the most reliable updates corresponding to large derivatives in the Fréchet matrix (e.g., those in the cells crossed by dense VSP rays).

To improve the convergence of the algorithm, we propose a three-stage parameter-updating procedure. In the first several iterations, only the velocity \( V_{p0} \) is updated on a grid, while the anisotropy parameters \( \epsilon \) and \( \delta \) are fixed at their initial values. This operation eliminates potentially large distortions in \( \epsilon \) and \( \delta \) caused by the parameter trade-offs. At the second stage of the inversion, \( \epsilon \) and \( \delta \) are taken constant in each layer and updated together with the grid-based velocity \( V_{p0} \). Finally, all three TTI parameters are estimated simultaneously on the grid with the constraints provided by the regularization terms described above.

First, the algorithm was tested on a model that contains a bending TTI thrust sheet composed of several dipping blocks. Because of the relatively simple model geometry, we applied only the second stage of the updating procedure by estimating the velocity \( V_{p0} \) on a grid, while keeping \( \epsilon \) and \( \delta \) constant in each layer. This “quasi-factorized” assumption proved sufficient to recover both \( \epsilon \) and \( \delta \) due to a wide range of available reflector dips. Without walkaway VSP traveltimes or strong regularization, however, the velocity \( V_{p0} \) defined on a small grid could not be resolved just by flattening the CIGs for the available three reflectors.

Then the joint tomography of reflection and VSP data with structure-guided regularization was successfully applied to two sections of the BP TTI model that include an anticline and a salt dome. In both tests, a purely isotropic velocity field, which was obtained from check-shot traveltimes and extrapolated along the horizons, served as the initial model. With constraints from P-wave reflection and VSP data, the TTI parameters in the shallow part (above 5 km) of both sections are well-resolved. However, the errors in the anisotropy parameters \( \epsilon \) and \( \delta \) increase with depth due to the small offset-to-depth ratio and poor coverage of VSP rays. For the model with the salt dome, the anisotropic velocity field is recovered with higher accuracy to the left of the salt, where the inversion was tightly constrained by VSP data from a nearby well.

5 ACKNOWLEDGMENTS

We would thank BP for providing the synthetic data set. This work was supported by the Consortium Project on Seismic Inverse Methods for Complex Structures at CWP.

REFERENCES


Joint migration velocity analysis of PP- and PS-waves for VTI media

Pengfei Cai & Ilya Tsvankin
Center for Wave Phenomena, Colorado School of Mines

ABSTRACT
Combining PP-waves with mode-converted PS-waves in migration velocity analysis (MVA) can help build more accurate VTI (transversely isotropic with a vertical symmetry axis) velocity models. To take advantage of efficient MVA algorithms designed for pure modes, here we generate pure SS-reflections from PP and PS data using the PP+PS=SS method. Then the residual moveout in both PP and SS common-image gathers is minimized during iterative velocity updates. The model is divided into square cells, and the VTI parameters $V_{P0}$, $V_{S0}$, $\epsilon$, and $\delta$ are defined at each grid point. The objective function also includes the differences between the migrated depths of the same reflectors on the PP and SS sections. The replacement of PS-waves with pure SS reflections in MVA allows us to avoid problems caused by the moveout asymmetry and other undesirable features of mode conversions. Synthetic examples confirm that 2D MVA of PP- and PS-waves can resolve all four relevant parameters of VTI media if reflectors with at least two distinct dips are available. After the velocity model has been reconstructed, accurate depth images can be obtained by migrating the recorded PP and PS data.

Key words: joint migration velocity analysis, PS-waves, VTI, codephiting

1 INTRODUCTION

Prestack depth migration (PSDM) and reflection tomography in the migrated domain are widely used in P-wave imaging (Stork, 1992; Wang et al., 1995; Adler et al., 2008; Bakulin et al., 2010). Most current PSDM and migration velocity analysis (MVA) algorithms account for transverse isotropy with a vertical (VTI) or tilted (TTI) symmetry axis. Sarkar and Tsvankin (2004) develop an efficient MVA method for VTI media by dividing the model into factorized blocks. Within each block, the anisotropy parameters are constant, while the P-wave symmetry-direction velocity $V_{P0}$ varies linearly in space. To extend MVA to more complex subsurface structures, Wang and Tsvankin (2010) suggest a P-wave ray-based gridded tomographic algorithm in which the parameters $V_{P0}$, $\epsilon$, and $\delta$ are defined on a rectangular grid. Similar multiparameter tomographic inversion is applied to synthetic and field P-wave data by Zhou et al. (2011).

To resolve the velocity $V_{P0}$ and anisotropy parameters $\epsilon$ and $\delta$ required for P-wave depth imaging, it is necessary to combine P-wave traveltimes with additional information. Tsvankin and Thomsen (1995) demonstrate that long-spread (nonhyperbolic) P- and SV-wave moveouts are sufficient for estimating the VTI parameters $V_{P0}$, $\epsilon$, $\delta$, and the SV-wave vertical velocity $V_{S0}$. However, it is more practical to supplement P-waves with converted PS(PSV) data. For 2D VTI models, the parameters $V_{P0}$, $V_{S0}$, $\epsilon$, and $\delta$ can be obtained by combining PP- and PS-wave traveltimes for a horizontal and dipping interface (Tsvankin and Grechka, 2000).

Several authors discuss joint tomographic inversion of P and PS data (Stopin and Ehinger, 2001; Audebert et al., 1999; Broto et al., 2003; Foss et al., 2005). However, velocity analysis of mode conversions is hampered by the asymmetry of PS moveout (i.e. PS traveltimes generally do not stay the same when the source and receiver are interchanged) and polarity reversals of PS-waves. As discussed by Thomsen (1999) and Tsvankin and Grechka (2011), the apex of the PS moveout in common-midpoint (CMP) gathers typically is shifted from zero offset. Therefore, MVA for PS-waves (Du et al., 2012; Foss et al., 2005; Audebert et al., 1999) has to account for the “diodic” nature of PS reflections (Thomsen, 1999). For example, common-image gathers (CIGs) of PS-waves can be computed separately for positive and negative offsets in the tomographic objective function (Foss et al., 2005).

To replace mode conversions in velocity analysis with pure SS reflections, Grechka and Tsvankin (2002) suggest the so-called PP+PS=SS method. By combining PP and PS events that share P-legs, that method generates SS reflection data with the correct kinematics. Grechka et al. (2002b) perform joint inversion of PP and PS reflection data using so-called
stacking-velocity tomography, which operates with NMO velocities for 2D lines and NMO ellipses for wide-azimuth 3D surveys (Grechka et al., 2002a). They construct SS traveltimes with the PP+PS=SS method and then apply stacking-velocity tomography to the PP and SS data. However, their methodology is limited to hyperbolic moveout and excludes information contained in long-offset traveltimes (Tsvankin and Grechka, 2011). Also, stacking-velocity tomography can be applied only to relatively simple layered or blocked models.

To make use of the efficient MVA techniques developed for pure modes (Sarkar and Tsvankin, 2004; Wang and Tsvankin, 2010), here we apply the PP+PS=SS method to construct pure SS-wave reflections from PP and PS data. The MVA is performed by minimizing residual moveout of reflection events in both PP- and SS-wave CIGs. Our velocity-updating for P-wave data is based on the ray-based gridded tomography developed by Wang and Tsvankin (2010). PP and PS images of the same reflector do not match in depth if the velocity model is incorrect (Foss et al., 2005; Tsvankin and Grechka, 2011). Therefore, in addition to flattening image gathers, we penalize depth misties between PP and PS sections.

First, we introduce the methodology of joint MVA of PP- and PS-waves, which includes identification (registration) of PP and PS events from the same interface, application of the PP+PS=SS method to create SS reflection data, and joint MVA of the recorded PP- and generated SS-waves. Next, we test the algorithm on a simple model comprised of a VTI layer sandwiched between isotropic media and on a layered model that includes a dipping reflector (e.g., a fault plane). The testing shows that for the second model the joint MVA algorithm can resolve the VTI parameters without additional information.

2 METHODOLOGY

P-wave reflection traveltimes generally are insufficient to resolve the parameters \( V_{P0} \), \( \epsilon \), and \( \delta \) required for P-wave depth imaging in VTI media. Therefore, to build a VTI model for prestack depth migration, at least one medium parameter (e.g., \( V_{P0} \)) must be known a priori (Sarkar and Tsvankin, 2004).

As discussed in Tsvankin and Grechka (2000), combining P-wave traveltimes with the moveout of PS-waves converted at a horizontal and dipping interface can help constrain the vertical P- and SV-wave velocities and the parameters \( \epsilon \) and \( \delta \). The most significant problem in PS-wave velocity analysis is the moveout asymmetry with respect to zero offset in CMP geometry (Tsvankin and Grechka, 2011). Unless the model is laterally homogeneous and has a horizontal symmetry plane, the PS traveltimes do not stay the same when the source and receiver are interchanged. Therefore, most migration velocity analysis methods designed for pure modes cannot be directly applied to converted waves. Here, we employ the PP+PS=SS method (Grechka and Tsvankin, 2002) to produce pure SS reflection events from the PP and PS reflections generated at the same interface.

Implementation of the PP+PS=SS method requires event registration, or identification of PP and PS reflections from the same interfaces. The main idea of the method is to combine PP and PS events that share the same P-legs. This is done by matching time slopes (horizontal slownesses) on common-receiver gathers of PP- and PS-waves (Figure 1). Then the traveltime of the constructed SS wave (sometimes called the “pseudo-S” arrival) is given by:

\[
t_{SS}(x_3,x_4) = t_{PS}(x_1,x_3) + t_{PS}(x_2,x_4) - t_{PP}(x_1,x_2).
\]

While the constructed SS-wave traveltimes are exact, the PP+PS=SS cannot produce correct reflection amplitudes. Hence, we convolve the SS traveltimes with a Ricker wavelet to generate “pseudo” SS reflection data to be used for MVA. The maximum reflection angle of the shear wave generated by mode conversion in a horizontal isotropic layer with the P- and S-wave velocities \( V_P \) and \( V_S \) is \( \theta_{SS}^{crit} = \sin^{-1}(V_S/V_P) \), and the half-offset \( h_S \) cannot exceed the critical value,

\[
h_{SS}^{crit} = D \tan \left( \sin^{-1} \left( \frac{V_S}{V_P} \right) \right).
\]

D is the layer’s thickness. For example, for a typical \( V_S/V_P = 1/2 \), the maximum half-offset is less than 0.6D. Therefore, it is necessary to include long-offset PP and PS data to generate SS data suitable for robust velocity analysis. The offsets of the computed SS-wave are smaller than those for the acquired PP and PS data but may be sufficient for MVA if the survey includes offsets reaching two target depths.

Here, we extend the MVA algorithm of Wang and Tsvankin (2010) to multicomponent (PP and SS) data. The model is divided into square cells, and the parameters \( V_{P0}, V_{S0}, \epsilon, \) and \( \delta \) are defined at each grid point. We apply prestack Kirchhoff depth migration to both PP and SS data starting with initial isotropic velocity models. The moveouts of migrated PP and SS events in common-image gathers serve as input to the joint MVA. To constrain the parameters \( \eta \) and \( \epsilon \), the moveout in PP CIGs is described by the nonhyperbolic equation (Sarkar

![Figure 1. Matching the horizontal slownesses on common-receiver PP and PS sections at locations \( x_1 \) and \( x_2 \) helps find the source-receiver coordinates \( x_3 \) and \( x_4 \) of the pure SS ray \( x_3Rx_4 \). This reconstructed SS ray has the same reflection point \( R \) as the PP ray \( x_1Rx_2 \) and PS rays \( x_1Rx_3 \) and \( x_2Rx_4 \) (after Grechka and Tsvankin, 2002).](image-url)
and Tsvankin, 2004):

\[ z^2(h) = z^2(0) + Ah^2 + B \frac{h^4}{h^2 + z^2(0)}, \quad (3) \]

where \( z \) is the migrated depth, and the coefficients \( A \) and \( B \) are found by a 2D semblance scan. There is no need to apply equation 3 to SS-wave CIGs because the offset-to-depth ratio of the constructed SS events seldom exceeds 1-1.2. For the joint MVA, we not only minimize the residual moveout in PP- and SS-wave CIGs, but also perform codepthing, which involves tying PP and SS images of the same reflectors. The objective function includes a term that penalizes the mismatch in depth of PP and SS migrated images using a selection of key reflection points. Those points are chosen on the basis of coherency and focusing (Foss et al., 2005).

To update model parameters, it is necessary to compute traveltimes derivatives with respect to the model parameters (Wang and Tsvankin, 2010). Since we operate with PP- and SS-waves, the parameter set includes not just \( V_{P0}, \epsilon, \) and \( \delta \), but also the shear-wave vertical velocity \( V_{S0} \). The exact \( P \)- and SV-wave phase velocities in VTI media can be expressed as (Tsvankin, 2005):

\[
\frac{V_{P0}^2}{V_{S0}^2} = 1 + \epsilon \sin^2 \theta - \frac{f}{2}
\]

\[ \pm \frac{f}{2} \sqrt{1 + 4 \sin^2 \theta \left( \frac{f}{2} \right) (2 \delta \cos^2 \theta - \epsilon \cos 2\theta) + \frac{4 \epsilon^2 \sin^4 \theta}{f^2}} \]

where \( \theta \) is the phase angle with the symmetry axis and \( f \equiv 1 - \frac{V_{S0}^2}{V_{P0}^2} \). The plus in front of the radical corresponds to \( P \)-waves and minus to \( S \)-waves. For purposes of MVA, however, it is convenient to replace \( \epsilon \) and \( \delta \) with the P-wave horizontal \( (V_{\text{hor},P}) \) and NMO \( (V_{\text{nmo},P}) \) velocities given by \( V_{\text{hor},P} = V_{P0} \sqrt{1 + 2\epsilon} \) and \( V_{\text{nmo},P} = V_{P0} \sqrt{1 + 2\delta} \). We compute the traveltime derivatives with respect to \( V_{\text{hor},P} \) and \( V_{\text{nmo},P} \) instead of \( \epsilon \) and \( \delta \). The MVA algorithm updates \( V_{P0}, \) \( V_{S0}, \) \( V_{\text{hor},P}, \) and \( V_{\text{nmo},P} \) and then converts \( V_{\text{hor},P} \) and \( V_{\text{nmo},P} \) to \( \epsilon \) and \( \delta \).

The objective function used in the joint MVA is as follows:

\[
F(\Delta \lambda) = \mu_1 ||A_P \Delta \lambda + b_P||^2 + \mu_2 ||A_S \Delta \lambda + b_S||^2 + \mu_3 ||D \Delta \lambda + y||^2 + \zeta ||L \Delta \lambda||^2,
\]

(5)

where \( A_P \) and \( A_S \) depend on the derivatives of the PP and SS migrated depths with respect to the medium parameters, the vectors \( b_P \) and \( b_S \) contain elements that characterize the residual moveout in PP- and SS-wave CIGs, the matrix \( D \) describes the differences between the derivatives of the PP and SS migrated depths with respect to the medium parameters, and the vector \( y \) contains the differences between the migrated depths on the PP and SS sections. The full definitions of \( A_P, A_S, D, b_P, b_S, \) and \( y \) are given in Appendix A. Minimizing the first two terms (\( ||A_P \Delta \lambda + b_P||^2 \) and \( ||A_S \Delta \lambda + b_S||^2 \)) allows us to flatten PP and SS CIGs. Codepthing is achieved through minimizing the third term (\( ||D \Delta \lambda + y||^2 \)). Because the tomographic inversion can be ill-posed, we add a regularization term (\( ||L \Delta \lambda|| \)) to the objective function. The coefficients \( \mu_1, \mu_2, \mu_3, \) and \( \zeta \) govern the weights of the corresponding terms. The objective function is minimized by a least-squares algorithm.

3 TESTS ON SYNTHETIC DATA

We use anisotropic ray-tracing package ANRAY to generate PP- and PS-wave reflection traveltimes for synthetic tests. ANRAY is developed by the consortium project “Seismic Waves in Complex 3D Structures” (SW3D) at Charles University in Prague. PP and SS images are generated with Kirchhoff prestack depth migration (Seismic Unix program ‘sukdepth3’). To create traveltime tables of PP- and SS-waves, we perform ray tracing using SU code ‘ray2dan’.

We first test the algorithm on a simple horizontally layered model (Figure 3) that includes a VTI layer sandwiched between isotropic media. Without dipping interfaces, the interval parameters for this model cannot be constrained by PP and PS reflection traveltimes. Therefore, the P-wave vertical
velocity $V_{P0}$ is assumed to be known, and we invert only for $V_{S0}$ and the anisotropy parameters $\epsilon$ and $\delta$ of the VTI layer.

The top layer is known to be isotropic and its P- and S-wave velocities can be easily computed from reflection data. The initial S-wave vertical velocity for the middle layer has an error of 10%, and both $\epsilon$ and $\delta$ are set to zero. The maximum offset-to-depth ratio for the bottom of the VTI layer is close to two, which is sufficient for applying the PP+PS=SS method. Indeed, the maximum offset for recorded PP data is 3 km and the maximum offset for constructed SS data is 1.6 km. The whole PP data set is used along with the SS-waves to estimate the residual moveout in CIGs. For codepthing, however, we only use conventional-spread PP data with offsets not exceeding those for SS-waves.

Figures 4(a) and 4(b) show the PP- and SS-wave migrated sections obtained with the initial model. The bottom of the VTI layer is poorly focused and is imaged at different depths on the PP and SS sections; also, the PP and SS image gathers are not flat (Figures 5(a) and 5(b)).

CIGs used for velocity analysis are uniformly sampled from 1 to 3 km along the second reflector. Since we use gridded tomography, the derivatives of migrated depths with respect to the model parameters are calculated at the vertices of relatively fine grids. Therefore, we follow Wang and Tsvankin (2010) in employing a mapping matrix to convert the model updates into the parameter values at each grid point. After 10 iterations, the CIGs are flat (Figures 6(a) and 6(b)) and the reflectors are tied in depth (Figures 7(a) and 7(b)). The estimated parameters of the VTI layer are close to the actual values: $V_{S0} = 1494$ m/s, $\epsilon = 0.1$, and $\delta = -0.09$.

Next, the algorithm is tested on a model (Figure 8) that includes a reflector (e.g., a fault plane) dipping at an angle approaching 30°. To avoid instability in ray tracing, we smooth the corner of the dipping interface using bicubic spline interpolation. The synthetic data include PP and PS reflections from both horizontal and dipping reflectors (Figure 9). The maximum offsets are 4 km for PP data and 2 km for SS-waves constructed by the PP+PS=SS method. Tsvankin and Grechka (2000) demonstrate that the traveltimes of the PP- and PS-waves reflected from a horizontal and a dipping interface are sufficient to constrain the parameters $V_{P0}$, $V_{S0}$, $\epsilon$, and $\delta$.

Here we only invert for the parameters of the middle layer. The initial model is isotropic with the velocities $V_{P0}$ and $V_{S0}$ distorted by 15%. The depth sections and CIGs computed with the initial model are displayed in Figures 10 and 11. A set of CIGs of PP- and SS-waves from both the horizontal and
disjoint MVA of PP- and PS-waves

Figure 8. Three-layer VTI model with dipping interfaces. The parameters of the first layer are $V_P = 2000$ m/s, $V_S = 1000$ m/s; for the second layer, $V_{P0} = 3000$ m/s, $V_{S0} = 1500$ m/s, $\epsilon = 0.2$ and $\delta = 0.1$. The maximum dip of both reflectors is 27°.

Figure 9. CMP gathers of the recorded (a) PP-waves and (b) PS-waves at location 3000 m. (c) The SS data constructed by the PP+PS=SS method at the same location.

Figure 10. (a) PP-wave and (b) SS-wave depth images computed with the initial model parameters.

Figure 11. Common-image gathers of (a) PP-waves and (b) SS-waves (displayed every 100m) after migration with the initial model.

rate estimates of the interval VTI parameters: $V_{P0} = 3031$ m/s, $V_{S0} = 1515$ m/s, $\epsilon = 0.20$, and $\delta = 0.08$.

4 DISCUSSION AND CONCLUSIONS

In the presence of moderate dips, combining PP reflection traveltimes with converted PS data may help reconstruct VTI velocity models in depth. Here, we presented an efficient algorithm for joint migration velocity analysis of PP- and PS-waves from heterogeneous VTI media.

To avoid problems caused by the moveout asymmetry and other inherent features of mode conversions, we construct pure SS reflections using the PP+PS=SS method. Then migration velocity analysis is performed for PP and SS data, which allows us to employ existing tomographic techniques designed for pure modes. In addition to flattening common-image gathers of PP- and SS-waves, the joint MVA is designed to remove the depth misties between PP and SS sections. After performing prestack Kirchhoff depth migration on PP and SS data, we use a nonhyperbolic semblance scan to evaluate the long-spread residual moveout in CIGs. Then the velocity model is
updated iteratively by flattening PP- and SS-wave CIGs and tying of the PP and SS migrated images in depth.

Synthetic testing confirmed that if both horizontal and moderately dipping PP and PS events are available, the joint MVA converges toward the correct VTI depth model. Therefore, PS-waves can play an important role in velocity model building for prestack depth migration. Multicomponent data also provide accurate estimates of shear-wave vertical velocity $V_{S0}$, which can be used in lithology prediction and reservoir characterization.

5 ACKNOWLEDGMENTS

We are grateful to James Gaiser (Geokinetics) for helpful discussions and to the members of the Anisotropy-Team of the Center for Wave Phenomena (CWP) for valuable technical assistance.

APPENDIX A: PARAMETER UPDATING METHODOLOGY

We extend the MVA algorithm of Wang and Tsvankin (2010) to the combination of PP and SS data, where the SS-waves are constructed by the PP+PS=SS method. In addition to flattening PP and SS image gathers, we also perform codephasing to ensure that the reflectors on the PP and SS sections are imaged at the same depth. Following Wang and Tsvankin (2010), the variance $Var$ of the migrated depths can be written as

$$Var = \sum_{j=1}^{U} \sum_{k=1}^{M} [z(x_j, h_k) - \bar{z}(x_j)]^2,$$  \hspace{1cm} (A1)

where $U$ is the number of image gathers used in the velocity update, $M$ is the number of offsets in image gathers, $x_j$ is the midpoint, $h_k$ is the half-offset, and $\bar{z}(x_j) = (1/M) \sum_{k=1}^{M} z(x_j, h_k)$ is the average migrated depth at $x_j$. By minimizing the variances $Var_P$ (for P-waves) and $Var_S$ (for S-waves), we flatten CIGs for both modes. The difference between the migrated depths of PP- and SS-waves from the same reflector can be estimated as:

$$Var_{P-S} = \sum_{j=1}^{U} [z_p(x_j) - z_s(x_j)]^2.$$  \hspace{1cm} (A2)

Minimizing $Var_{P-S}$ makes it possible to tie the PP and SS sections in depth. Differentiating the variances with respect to the parameter updates $\Delta \lambda$ and setting $\partial V / \partial (\Delta \lambda) = 0$ yields

$$A_P^T A_P \Delta \lambda = -A_P^T b_P$$  \hspace{1cm} (A3)

$$A_S^T A_S \Delta \lambda = -A_S^T b_S$$  \hspace{1cm} (A4)

$$D^T \Delta \lambda = -D^T y,$$  \hspace{1cm} (A5)

where $A$ is a $M \times U$ by $W \times N$ matrix with elements $g_{jk,ic} - \tilde{g}_{jk,ic}$, where $g_{jk,ic}$ is $\partial z(x_j, h_k)/\partial \lambda_{ic}$ and $\tilde{g}_{jk,ic} = (1/M) \sum_{k=1}^{M} g_{jk,ic}$. The subscripts P and S denote P- and S-waves, respectively. The vector $b$ contains $M \times U$ elements defined as $z(x_j, h_k) - \bar{z}(x_j)$. $D$ is also a $M \times U$ by $W \times N$ matrix and its elements are $g_{jk,ic}^P - g_{jk,ic}^S$, where $g_{jk,ic}^P$ is computed for P-waves and $g_{jk,ic}^S$ for S-waves. The $M \times U$ vector $y$ has elements $z^P(x_j, h_k) - z^S(x_j, h_k)$ as discussed in Wang and Tsvankin (2010), the derivatives $g_{jk,ic}$ are found by differentiating phase velocity with respect to the medium parameters along the raypath.
REFERENCES


Foss, S., B. Ursin, and M. V. de Hoop, 2005, Depth-consistent reflection tomography using PP and PS seismic data: Geophysics, 70, U51–U65.


Thomsen, L., 1999, Converted-wave reflection seismology over inhomogeneous, anisotropic media: Geophysics, 64, 678–690.


Increasing illumination and sensitivity of reverse-time migration with internal multiples

Clement Fleury
Center for Wave Phenomena, Colorado School of Mines

ABSTRACT
Reverse-time migration is a two-way time-domain finite-frequency technique that accurately handles the propagation of complex scattered waves and produces a band-limited representation of the subsurface structure that is conventionally assumed to be linear in the model parameters. Because of this underlying linear single-scattering assumption, most implementations of this method do not satisfy energy conservation and do not optimally use illumination and model sensitivity of multiply scattered waves. Migrating multiply scattered waves requires preserving the nonlinear relation between image and model parameters. I modify the extrapolation of source and receiver wavefields to more accurately handle multiply scattered waves. I extend the concept of the imaging condition in order to map into the subsurface structurally coherent seismic events that correspond to the interaction of both singly and multiply scattered waves. This results in an imaging process here referred to as nonlinear reverse-time migration. It includes a strategy that analyzes separated contributions of singly and multiply scattered waves to the final nonlinear image. The goal is to provide a tool suitable for seismic interpretation and potentially migration velocity analysis that benefits from increased illumination and sensitivity from multiply scattered seismic waves. It is noteworthy that this method applies to migrating internal multiples, a clear advantage for imaging challenging complex subsurface features, e.g., in salt and basalt environments. The results of synthetic seismic imaging experiments, one of which includes a subsalt imaging example, illustrate the technique.

Key words: Imaging, Interpretation, Seismics, Velocity analysis, Multiples

1 INTRODUCTION
Complex subsurface structures, such as sub-basalt and subsalt structures, generate strong scattering which makes them challenging to image with conventional techniques (e.g., Martini and Bean, 2002; Leveille et al., 2011). For such geologic environments, multiply scattered waves, such as internal and surface multiples, contain useful information about the subsurface but these waves are traditionally suppressed, e.g., with multiple suppression (Foster and Mosher, 1992; ten Kroode, 2002) or surface-related multiple elimination (Verschuur et al., 1992; Dragošet et al., 2010). Alternatively, multiply scattered waves are usually just ignored, e.g., with migration of primaries (Claerbout, 1985; Esmeroy and Oristaglio, 1988), least-square migration (Nemeth et al., 1999) or linearized inversion (Symes, 2008a), when linearizing the relation between model and data during the imaging/inversion process. The motivations for using multiply scattered waves and their nonlinear relation to the model are to preserve amplitudes, provide extra illumination, account for energy conservation, and increase redundancy and sensitivity to model parameters.

Reverse-time migration (RTM) is now a standard imaging technique in the industry (Baysal et al., 1983; McMechan, 1989). The method provides high-quality images for oil and gas exploration and is of special interest for complex subsurface geologies (e.g., Farmer et al., 2006; Leveille et al., 2011). Yet, conventional RTM migration does not resolve the problem of imaging waves multiply scattered by complex subsurface structures. The underlying single-scattering assumption of most migration techniques does not account for the fundamental nonlinear relation between model and multiply scattered data (Fleury and Vasconcelos, 2012). To utilize the energy, illumination and sensitivity contained
in multiply scattered data, Fleury and Snieder (2011, 2012) have proposed a nonlinear RTM migration algorithm has been proposed. In this paper, I expand the description of this technique and define a clear strategy for using multiply scattered waves in seismic interpretation and migration velocity analysis.

Advances have been made in using multiples to perform imaging (Brown and Guitton, 2005; Jiang et al., 2007; Malcolm et al., 2009; Verschuur and Berkhourt, 2011). These methods apply redatuming techniques (Schuster et al., 2004; Malcolm and de Hoop, 2005; Berkhourt and Verschuur, 2006) for extrapolating seismic data and reconstructing scattered wavefields with kinematically correct multiples, but they do not modify the imaging condition to account for the nonlinearity of the resulting wavefields with respect to the model. As with these methods, the method presented in this document takes into account multiple scattered waves in the extrapolation procedure. It takes advantage of computationally affordable reverse-time migration engines to handle complex scattering wave phenomena and adapts the extrapolation procedure to include sources of scattering in the migration velocity model, which improves the reconstruction of scattered waves (Vasconcelos et al., 2010). Unlike the previous methods, the method presented here also extends the concept of the imaging condition in order to map into the subsurface structurally coherent seismic events that correspond to reflections of both singly and multiply scattered waves. In that sense, this method is in agreement with Verschuur and Berkhourt (2011) who call for a new imaging principle, defined as a minimization problem in their analysis. Fully exploiting multiple-scattering interactions results in images of multiples as well as primaries. The imaging condition here is nonlinear in the model parameters (Fleury and Vasconcelos, 2012) and produces four sub-images that I analyze as separate contributions to the final nonlinear image. This strategy leads to a new tool suitable for seismic interpretation and with the potential to improve migration velocity analysis.

First, I present the fundamentals of my nonlinear reverse-time migration (NLRTM) method and then show how this method leads to increased illumination and sensitivity for reverse-time migration with multiply scattered waves. Finally, we see the efficacy of this new seismic interpretation tool with application to synthetic data, including a subsalt imaging example.

2 THEORY OF NONLINEAR REVERSE-TIME MIGRATION

2.1 Wavefield extrapolation

The division of Earth properties into subdomains of short and long wavelengths justifies the use of a smooth model \( m_0 \) perturbed by a rough model \( \Delta m \) in seismic migration (Jannane et al., 1989). In the acoustic assumption, the model is given by \( m = m_0 + \Delta m = (\rho, c) \), where \( \rho \) and \( c \) are density and velocity, respectively. Model \( m_0 \) provides a kinematically correct wavefield \( w_0 \) and is used as a conventional migration velocity model. Model perturbation \( \Delta m \) acts as a scattering source for the scattered wavefield \( w_s \) superimposed on reference wavefield \( w_0 \). Model \( \Delta m \) is the part of the Earth model that one tries to represent with the seismic image \( i \).

For each shot of an active seismic survey, reference wavefield \( w_0 \) is obtained by propagating the estimated source signature \( s \) in migration model \( m_0 \):

\[
H_0 \cdot w_0 = s, \tag{1}
\]

where the acoustic wave operator is given by \( H_0 = \frac{1}{\rho_0} \nabla \cdot \left( \frac{1}{\rho_0} \nabla \right) + \frac{1}{c_0^2} \frac{\partial^2}{\partial t^2} \). The dot symbol \( \cdot \) denotes the tensor contraction between matrix operators and vector fields throughout this paper. Because model \( \Delta m \) is unknown in the imaging process, scattered wavefield \( w_s \) is not obtained by forward modeling but approximated by the receiver-side extrapolated wavefield \( w_{s,rec} \) which is conventionally obtained by backpropagating the recorded scattered data \( d_s \) in background model \( m_0 \),

\[
H_0^\dagger \cdot w_{s,rec} = d_s, \tag{2}
\]

where the adjoint acoustic wave operator is given by \( H_0^\dagger = \frac{1}{\rho_0} \nabla \cdot (\rho_0 \nabla) + \frac{1}{c_0^2} \frac{\partial^2}{\partial t^2} \).

The perturbation of operator \( H_0 \) describing scattering caused by model \( \Delta m \) is denoted by the operator \( P \) and is a function of models \( m_0 \) and \( \Delta m \) (e.g., Kato, 1995). The perturbed operator is defined as \( H_0 - P \). The relationship between wavefield \( w_s \) (equal to the scattered seismic data \( d_s \) at the receiver locations) and perturbation \( P \) is nonlinear (e.g., Weglein et al., 2003):

\[
w_s = \frac{1}{s} (w_0 \cdot P \cdot w_0) + \frac{1}{s^2} (w_0 \cdot P \cdot w_0 \cdot P \cdot w_0) + ... \tag{3}
\]

This nonlinear relationship between wavefield and perturbation fundamentally explains multiply scattered waves, i.e., waves that have been reflected, diffracted, and more generally scattered more than once, such as internal multiples. To preserve the nonlinear relation between wavefield and perturbation, I modify conventional RTM extrapolation. Given an estimate \( P_{est} \) of the perturbation operation, the scattered wavefield \( w_{s,rec} \) is extrapolated with a more exact equation than equation (2) (Vasconcelos et al., 2010):

\[
(H_0^\dagger - P_{est}^\dagger) \cdot w_{s,rec} = d_s + P_{est}^\dagger \cdot w_0. \tag{4}
\]

This is equivalent to including scattering sources in the extrapolation of scattered wavefield \( w_{s,rec} \) and results in better reconstruction of scattered waves. Additionally, let us introduce the source-side extrapolated scattered
wavefield $w_{s,sou}$ defined in the following relationship.

$$ (H_0 - P_{est}) \cdot w_{s,sou} = P_{est} \cdot w_0. \quad (5) $$

Wavefield $w_{s,sou}$ is originally missing in conventional RTM migration. In the conventional extrapolation procedure, this wavefield would be zero.

The introduction of operator $P_{est}$ in the extrapolation allows us to account for scattering contrasts in reconstructing scattered wavefields $w_{s,sou}$ and $w_{s,rec}$. For acoustic media, operator $P_{est}$ is defined as

$$ P_{est} = \rho_0 \nabla \cdot \frac{1}{\rho_0} \nabla - (\rho_0 + \Delta \rho_{est}) \nabla \cdot \left( \frac{1}{\rho_0 + \Delta \rho_{est}} \nabla \right) $$

$$ + \left( \frac{1}{c_0^2} - \frac{1}{(c_0 + \Delta c_{est})^2} \right) \frac{\partial^2}{\partial t^2}, \quad (6) $$

where model $\Delta m_{est}$ is an estimate of the true scattering contrast model $\Delta m$. There are two main approaches to retrieving model $\Delta m_{est}$. An estimate of model $\Delta m$ might possibly be obtained in an automatic/semi-automatic algorithm-based manner by essentially solving an inverse problem for model $\Delta m$ in the data domain (e.g., Tarantola, 1984, 1986). The linearization of this inverse problem, written as

$$ d_s = w_0 \cdot P_{est} \cdot w_0, \quad (7) $$

leads to a solution $\Delta m_{est}$ that can be computed by least-squares migration (Nemeth et al., 1999; Plessix and Mulder, 2004) or optimal scaling of conventional RTM images (Rickett, 2003; Symes, 2008a). The application of gradient-based methods allows for updating model $\Delta m_{est}$ in several linear iterations (e.g., Pratt et al., 1998). Therefore, the NLRTM method is fully integrable with full waveform inversion (FWI) technology (e.g., Tarantola, 1984; Pratt, 1999; Plessix, 2006; Virieux and Operto, 2009; Zhu et al., 2009). Alternatively, tools in velocity model building based on the input of a human interpreter could be used for creating a model with sharp interfaces. This interpreted model defines scattering model $\Delta m_{est}$ for the extrapolation of scattered wavefields $w_{h,sou}$ and $w_{h,rec}$. The next section contains examples using both approaches.

### 2.2 Imaging condition

The imaging condition maps into the subsurface the interaction of source-side extrapolated wavefields $w_0$ and $w_{s,sou}$ with receiver-side extrapolated wavefield $w_{s,rec}$ to create a representation of the Earth’s physical properties. Following the method of Fleury and Vasconcelos (2012), let us define image $i$ as

$$ i = \sum_{sources} w_0 \ast w_{s,rec} + \frac{1}{2} \sum_{sources} w_{s,sou} \ast w_{s,rec}, \quad (8) $$

where $\ast$ denotes either zero-time crosscorrelation or deconvolution. A crosscorrelation imaging condition corresponds to an energy-based representation of the subsurface and maps an estimate of “energy loss” in scattering (Fleury and Vasconcelos, 2012). A deconvolution imaging condition provides a reflectivity-based representation and maps into the subsurface a quantity that approximates reflectivity or scattering amplitude depending on the particular definitions of the imaging condition (e.g., de Bruin et al., 1990). In this paper, I adopt a deconvolution imaging condition following the recommendations of Schleicher et al. (2008).

This imaging condition is nonlinear with respect to perturbation $P$ and therefore with respect to model $\Delta m$ (Fleury and Vasconcelos, 2012), but preserves a linear relation between image $i$ and the data $d_s$. To emphasize these properties, let us separately analyze the two contributions, images $i_0$ and $i_s$, to image $i$ (properties summarized in Table 1). Image $i_0$ maps the interaction of reference wavefield $w_0$ with scattered wavefield $w_{s,rec}$. The imaging condition that defines image $i_0$,

$$ i_0 = \sum_{sources} w_0 \ast w_{s,rec}, \quad (9) $$

is similar to conventional imaging conditions for RTM migration. The only difference from the conventional imaging condition comes from the modified extrapolation of wavefield $w_{s,rec}$. Image $i_0$ is a function of perturbation $P$ and the data $d_s$ because of its dependency on scattered wavefield $w_{s,rec}$. Wavefield $w_{s,rec}$ is a linear function of the data $d_s$. As a result, image $i_0$ is linear with respect to the data $d_s$. Wavefield $w_{s,rec}$ is in general a nonlinear function of perturbation $P$. Only conventional reverse-time extrapolation under a linear single-scattering assumption for the scattered data $d_s$ gives a linear estimated wavefield $w_{s,rec}$ with respect to perturbation $P$. Image $i_0$ thus results from the linearization of equation (8) with respect to perturbation $P$ under a linear single-scattering assumption. This assumption is, however, not always accurate enough and more importantly limits the capability of RTM migration to image complex scattering events such as those that involve internal multiples (e.g., Malcolm et al., 2007, 2011). Image $i_s$ maps the interaction of scattered wavefields $w_{s,sou}$ and $w_{s,rec}$. The imaging condition that defines image $i_s$,

$$ i_s = \sum_{sources} w_{s,sou} \ast w_{s,rec}, \quad (10) $$

takes into account higher orders of scattering neglected under a linear single-scattering assumption. Through its dependencies of scattered wavefield $w_{s,rec}$, image $i_s$ is a function of the data $d_s$, and, similar to image $i_0$, image $i_s$ is linear with respect to the data $d_s$. Image $i_s$ is a function of perturbation $P$ because of its dependencies of both scattered wavefields $w_{s,sou}$ and $w_{s,rec}$. Wavefields $w_{s,sou}$ and $w_{s,rec}$ are at least linear in perturbation $P$ and are in general nonlinear in perturbation $P$. Image $i_s$ is therefore at least quadratic in pertur-
bution \( \mathbf{P} \) and is in general nonlinear in perturbation \( \mathbf{P} \). While image \( i_0 \) is therefore negligible under a linear single-scattering assumption, it is essential when imaging general complex scattering subsurface structures.

Years of application in industrial seismic exploration have shown that conventional RTM provides fairly accurate results. Apart from intrinsically problematic complicated structures (such as salt bodies and basalt formations) that break the assumption, linear single-scattering resolves most subsurface structures, or more precisely, it does so in the first order of approximation. In general, nonlinear scattering is not the dominant issue, but at the same time, it is rarely a negligible one. To first order in the perturbation \( \mathbf{P} \), image \( i_0 \) primarily contributes to image \( i_1 \), and image \( i_1 \) makes a secondary contribution. The examples in the next section demonstrate that image \( i_1 \) nonetheless provides amplitude compensation, complementary illumination, and increased sensitivity to the model parameters. To take advantage of these qualities of the NLRTM method, let us consider an analysis based on the separate and comparative interpretation of images \( i_0 \) and \( i_1 \). To emphasize only structural contrasts in the images, we decompose images \( i_0 \) and \( i_1 \) into

\[
i_0 = \sum_{\text{sources}} (w_0^{(d)} \ast w_{s,rec}^{(u)} + w_0^{(u)} \ast w_{s,rec}^{(d)})
\]

and

\[
i_1 = \sum_{\text{sources}} (w_{s,sou}^{(d)} \ast w_{s,rec}^{(u)} + w_{s,sou}^{(u)} \ast w_{s,rec}^{(d)})
\]

where superscripts \( (u) \) and \( (d) \) refer to up- and down-going wavefields \( w = w^{(d)} + w^{(u)} \), respectively, and redefine these images to keep only the interactions of up- and down-going wavefields:

\[
i_0 = \sum_{\text{sources}} w_0^{(d)} \ast w_{s,rec}^{(u)} + \sum_{\text{sources}} w_0^{(u)} \ast w_{s,rec}^{(d)}
\]

\[
i_1 = \sum_{\text{sources}} w_{s,sou}^{(d)} \ast w_{s,rec}^{(u)} + \sum_{\text{sources}} w_{s,sou}^{(u)} \ast w_{s,rec}^{(d)}
\]

The result is the four sub-images \( i_0^{(d,u)}, i_0^{(u,d)}, i_1^{(d,u)}, \) and \( i_1^{(u,d)} \) that contribute to nonlinear image \( i \). For example, \( i_0^{(d,u)} \) denotes the image of the interaction of down-going reference wavefield \( w_0^{(d)} \) and up-going scattered wavefield \( w_{s,rec}^{(u)} \). These four sub-images are key for our analysis. The up/down wavefield decomposition in the imaging condition isolates back-scattered (including reflected) from forward-scattered (including transmitted) energies. The redefinition of images \( i_0 \) and \( i_1 \) separates back-scattering from forward-scattering and reduces what has been referred to as low-frequency RTM artifacts or “transmission” artifacts (Liu et al., 2011; Fleury and Vasconcelos, 2012). The modified images \( i_0 \) and \( i_1 \) do not map exact scattering amplitudes (or energy loss in scattering for crosscorrelation-based imaging) because forward-scattering is ignored. The goal is to reveal subsurface scattering contrasts rather than to achieve true amplitude nonlinear scattering-based imaging, which, like conventional imaging, has some other intrinsic limitations, such as limited acquisition geometry, intrinsic attenuation, anisotropy, or shortcomings in multi-model parameter estimation (e.g., Gray, 1997; Deng and McMechan, 2007; Virieux and Operto, 2009).

The four NLRTM sub-images of equations (13) and (14) prove to be of particular interest for seismic interpretation and model sensitivity analysis.

### 3 STRATEGIES FOR NONLINEAR REVERSE-TIME MIGRATION

#### 3.1 Energy, illumination, and sensitivity of multiply scattered waves

The NLRTM method incorporates multiply scattered waves into the imaging process by taking the fundamental nonlinear relation between model and data into
Nonlinear reverse-time migration

Figure 1. Model $m_0 + \Delta m = (\rho_0 + \Delta \rho, c_0 + \Delta c)$ for experiments (a) point-scatterer velocity model, (b) high-velocity lens and point-scatterer velocity model, and (c) low-velocity zone and point scatterer velocity model. Two-layer density model (d) is the same for all three experiments. The red dots and green line indicate the fixed-spread source/receiver geometry.

Sub-images $i^{(d,u)}_d$ and $i^{(u,d)}_u$ primarily map single scattering events into the subsurface. These two sub-images mainly result from the illumination and sensitivity of singly scattered waves. Sub-images $i^{(d,u)}_s$ and $i^{(u,d)}_s$ map only multiple scattering events into the subsurface, and thus provide illumination and sensitivity information from multiply scattered waves. All sub-images are nonetheless representations of the same subsurface structure and must consequently share common structural features. A comparative study of the NLRTM sub-images therefore provides valuable information for interpreting subsurface structures. Additionally, the redundancy and consistency between NLRTM sub-images are new appreciable criteria for migration velocity analysis, which I briefly describe in this document and further exploit in future studies.

The three synthetic experiments in Figure 1 illustrate the fundamental properties of the NLRTM method. These experiments, denoted (a-c), emphasize different aspects of the use of NLRTM sub-images. I use the three models in Figure 1 to synthesize the seismic data and then migrate all the data with the same identical reference model (Figure 2). Experiment (a) corresponds to having a correct migration velocity model. Experiments (b) and (c) correspond to having an incorrect migration velocity model with either a strong localized anomaly (missing high-velocity lens) or a weak diffused anomaly (missing low-velocity zone), respectively. Despite these small velocity anomalies, the conventional RTM images in all three experiments provide an interpretable image of the density-constrast reflector (for example, see the RTM image in Figure 3 for experiment (a)). After estimating the focus quality at the point scatterer, one might question the reliability of the interpretation of the point scatterer especially for experiments (b) and (c). Based on this interpretation of the conventional RTM images, I create scattering model estimate $\Delta m_{est}$ (Figure 4), which contains the interpreted density reflector and leaves out the point scatterer. Model $\Delta m_{est}$ defines the perturbation operator $P_{est}$ used in the modified extrapolation procedure of the NLRTM account. The NLRTM sub-images result from mapping the entire linear and nonlinear scattered data into structurally coherent seismic events that correspond to the interaction of both singly and multiply scattered waves. The recorded multiply scattered data contain subsurface information that conventional RTM migration does not exploit: multiply scattered waves carry additional energy beyond that in singly scattered waves, illuminate the subsurface with better coverage than do singly scattered waves and are more sensitive to the Earth model than are singly scattered waves. In this section, I define a strategy for using energy, illumination, and sensitivity of multiply scattered waves in seismic interpretation. We shall also see that these attributes of multiply scattered waves are potentially useful for migration velocity analysis.

Sub-images $i^{(d,u)}_d$ and $i^{(u,d)}_u$ primarily map single scattering events into the subsurface. These two sub-images mainly result from the illumination and sensitivity of singly scattered waves. Sub-images $i^{(d,u)}_s$ and $i^{(u,d)}_s$ map only multiple scattering events into the subsurface, and thus provide illumination and sensitivity information from multiply scattered waves. All sub-images are nonetheless representations of the same subsurface structure and must consequently share common structural features. A comparative study of the NLRTM sub-images therefore provides valuable information for interpreting subsurface structures. Additionally, the redundancy and consistency between NLRTM sub-images are new appreciable criteria for migration velocity analysis, which I briefly describe in this document and further exploit in future studies.

The three synthetic experiments in Figure 1 illustrate the fundamental properties of the NLRTM method. These experiments, denoted (a-c), emphasize different aspects of the use of NLRTM sub-images. I use the three models in Figure 1 to synthesize the seismic data and then migrate all the data with the same identical reference model (Figure 2). Experiment (a) corresponds to having a correct migration velocity model. Experiments (b) and (c) correspond to having an incorrect migration velocity model with either a strong localized anomaly (missing high-velocity lens) or a weak diffused anomaly (missing low-velocity zone), respectively. Despite these small velocity anomalies, the conventional RTM images in all three experiments provide an interpretable image of the density-constrast reflector (for example, see the RTM image in Figure 3 for experiment (a)). After estimating the focus quality at the point scatterer, one might question the reliability of the interpretation of the point scatterer especially for experiments (b) and (c). Based on this interpretation of the conventional RTM images, I create scattering model estimate $\Delta m_{est}$ (Figure 4), which contains the interpreted density reflector and leaves out the point scatterer. Model $\Delta m_{est}$ defines the perturbation operator $P_{est}$ used in the modified extrapolation procedure of the NLRTM
method. For the three experiments, let us focus our attention on the reconstruction of the image of the point scatterer. Figures 5 to 7 show the NLRTM sub-images of the point scatterer (region of interest indicated by the square box of Figure 3) corresponding to experiments (a) through (c), respectively. In these figures, arrows provide a schematic representation of the direction of illumination resulting from up- and down-going reference and scattered wavefields. Sub-images $i_{(u,d)}^{(d,u)}$ and $i_{(d,u)}^{(u,d)}$ (shown in Figure 5 for experiment (a)) are discarded in the analysis because wavefields $w_{0}^{(u)}$ and $w_{s,sou}^{(d)}$, which contribute to sub-images $i_{(u,d)}^{(d,u)}$ and $i_{(d,u)}^{(u,d)}$, do not illuminate the point scatterer. These two sub-images are therefore identically zero at the point scatterer and do not contribute to its image.

For experiment (a), sub-images $i_{(d,u)}^{(d,u)}$ and $i_{(u,d)}^{(u,d)}$ (Figures 5a and 5b) show focusing of the scattered energy at the correct location of the point scatterer, that is at the location of the scatterer (indicated by the blue dot) with the tolerance of a small vertical variation due to the spatial extent of the point scatterer (modeled by a localized gaussian anomaly). With the correct velocity model, the NLRTM method accurately maps both singly and multiply scattered energy. This consistency between sub-images $i_{(d,u)}^{(d,u)}$ and $i_{(u,d)}^{(u,d)}$ is valuable information to ensure correct interpretation of the seismic image of the point scatterer. In sub-image $i_{0}^{(d,u)}$, reference wavefield $w_{0}^{(d)}$ illuminates the point scatterer from above, and, in sub-image $i_{0}^{(u,d)}$, scattered wavefield $w_{s,sou}^{(u)}$ illuminates the point scatterer from below. The NLRTM method provides extra illumination (represented by arrows in Figure 5) of the subsurface. Experiment (a) thus illustrates how to utilize energy and illumination of multiply scattered waves to provide extra
Nonlinear reverse-time migration

For experiment (b), the high-velocity lens causes the point scatterer to be focused at the incorrect location in depth in sub-image $i_{0}^{(d,u)}$ (Figure 6a) but does not misposition the reconstructed point scatterer in sub-image $i_{s}^{(u,d)}$ (Figure 6b) because wavefields $w_{s,rec}^{(u)}$ and $w_{s,sou}^{(d)}$ are sensitive to the high-velocity lens. For experiment (c), these observations are reversed. The velocity anomaly caused by the low-velocity zone does not affect sub-image $i_{0}^{(d,u)}$ (Figure 7a) but causes the multiply scattered energy by the point scatterer to defocus in sub-image $i_{s}^{(u,d)}$ (Figure 7b) because, contrary to wavefields $w_{0}^{(d)}$ and $w_{s,rec}^{(u)}$, wavefields $w_{s,sou}^{(u)}$ and $w_{s,rec}^{(d)}$ are sensitive to the low-velocity zone. With an incorrect velocity model, the NLRTM method does not simultaneously map singly and multiply scattered energy with accuracy. Comparing sub-images $i_{0}^{(d,u)}$ and $i_{s}^{(u,d)}$ thus provides a diagnostic for interpreting the reliability of the reconstruction of the point scatterer. For both experiments, sub-images $i_{0}^{(d,u)}$ and $i_{s}^{(u,d)}$ apply different illumination (represented by arrows in Figures 6 and 7) for mapping the point scatterer in depth. At the point scatterer location, the incident and scattered wavefields that contribute to either sub-images $i_{0}^{(d,u)}$ or $i_{s}^{(u,d)}$ propagate through different parts of the model and consequently do not exhibit the same sensitivity to the model. Experiments (b) and (c) demonstrate how illumination and sensitivity of multiply scattered waves control the quality of the interpretation of a seismic image. The gain in sensitivity that comes from the use of multiply scattered waves in the NLRTM method also provides valuable information to cross-validate the accuracy of a given migration velocity model. Discrepancies between NLRTM sub-images are potentially usable for velocity model building by setting an inverse problem to penal-
ize such discrepancies in the image space (e.g., de Hoop et al., 2006; Symes, 2008b, 2009) in a fashion similar to that in image-domain waveform tomography (e.g., Sava and Biondi, 2004; Shen and Symes, 2008; Yang and Sava, 2011).

### 3.2 Target-oriented subsalt imaging

A practical application for the NLRTM method is target-oriented subsalt imaging. Subsalt imaging is challenging because of the structural complexity of salt bodies, which leads to the general lack of illumination below salt (e.g., Muerdter and Ratcliff, 2001; Leveille et al., 2011). Energy, illumination, and sensitivity of multiply scattered waves are of potential interest for such a complex scattering geologic environment. The synthetic subsalt example of this subsection shows the efficacy of seismic interpretation based on NLRTM sub-images.

For the Sigsbee 2A model (Paffenholz et al., 2002), let us consider synthetic marine data generated using time-domain finite-difference modelling with the acquisition geometry and stratigraphic model shown in Figure 8. Figure 9 shows the migration velocity model, which contains interpreted hard salt boundaries. Conventional RTM migration provides the two images in Figure 10. A crosscorrelation-based imaging condition yields the image in Figure 10a. A Laplacian filter has been applied to the image in Figure 10a to reduce the low-frequency RTM artifacts commonly observed on the top of the salt body. After interpreting this image, let us notice poorly illuminated areas below the salt body and select the region of interest indicated by the green box in Figure 10a as the target region for improving the image quality and resolving ambiguities in this subsalt area after application of the NLRTM method. For comparison, Figure 10b is a conventional RTM image in the region of interest using a deconvolution-based imaging condition. I use a crosscorrelation-based imaging condition for the image in Figure 10a in order to build an estimate of perturbation model $\Delta m$ with an automatic algorithm-based method. Gradient-based data-domain misfit inversion methods, such as in FWI, require scaling of the gradient for model updating at each iteration (e.g., Pratt et al., 1998). Using a scaling function that is similar to the one used in FWI, I scale the image in Figure 10a to obtain model estimate $\Delta m_{est}$ (Figure 11) for use in the NLRTM method.

The NLRTM method produces the sub-images in Figure 12. Sub-image $i_0^{(d,u)}$ (Figure 12a) exhibits almost identical structure and only slightly different amplitude as compared to the conventional image (Figure 10b). The illumination is relatively poor below the
going scattered wavefield $w_{s,sou}^{(u)}$ with down-going scattered wavefield $w_{s,sou}^{(d)}$. The up-going reflected energy in wavefield $w_{s,sou}^{(u)}$ provides additional illumination and helps to improve the resolution of subsalt structures. Interestingly, structures that might be masked when illuminated from above might be observable from below. As shown in the previous sub-section, sub-image $i_{u,d}^{(u,d)}$ also exhibits sensitivity to the velocity model that is complementary to the sensitivity of the other three NLRTM sub-images. New and redundant structural information contained in sub-image $i_{u,d}^{(u,d)}$ facilitates more detailed seismic interpretation of the subsalt target. The coherent structures in sub-image $i_{u,d}^{(u,d)}$ are similar to those present in sub-images $i_{u,d}^{(d,u)}$ and $i_{u,d}^{(u,d)}$. This consistency is further evidence that migration velocity model $c_0$ (Figure 9) is accurate. Sub-image $i_{u,d}^{(u,d)}$ also shows better coherence of poorly illuminated reflectors and reveals additional features in the image. For example, the sediment layering across the two faults (shown as lines in Figure 12) is more visible. Comparative study of NLRTM sub-images thus helps in interpreting subsalt images in poorly illuminated areas.

**4 CONCLUSIONS**

The strategy for nonlinear reverse-time migration (NLRTM) here outlines the potentials of utilizing the energy, illumination, and sensitivity of multiply scattered waves in seismic imaging. The nonlinear relation between seismic model and data is key to incorporating multiply scattered waves into reverse-time migration. To take account of this nonlinear relation, the method estimates a scattering contrast model, modifies wavefield extrapolation, and extends the concept of the imaging condition. As a result, the NLRTM method outputs a panel of four nonlinear sub-images that represent the same subsurface structure from different aspects. It is possible to combine these four sub-images into a single nonlinear image of the subsurface, but this operation is nontrivial and must be adaptive to account for the amplitude variations across sub-images.

A comparative analysis of these nonlinear sub-images is a tool for both interpretation and possibly model sensitivity analysis. The target-oriented subsalt imaging example here illustrates the application of the NLRTM method: each NLRTM sub-image emphasizes different illumination of the subsurface structure, which results in additional information for the description of geological subsurface features. The consistency between redundant information across different sub-images gives a more confident assessment for seismic interpretation.

This same consistency criterion reveals the gain in sensitivity to model parameters that comes from the use of multiply scattered waves. In the future, I intend to use this extra sensitivity to develop new algorithms for migration velocity analysis. Current tools for migration
velocity analysis rely on assessing the quality of focus of conventional linear images. These techniques should be comparably applicable to the more sensitive nonlinear sub-images of the NLRTM method and should additionally benefit from the introduction of a new consistency criterion among sub-images.

5 ACKNOWLEDGMENTS

This work was supported by the Consortium Project on Seismic Methods for Complex Structures at the Center for Wave Phenomena. This research was supported in part by the Golden Energy Computing Organization at the Colorado School of Mines using resources acquired with financial assistance from the National Science Foundation and the National Renewable Energy Laboratory. I thank Professor Roel Snieder for his advice, support, and encouragement that contributed to the development of this research project. I am grateful to my CWP colleagues for their comments that helped improving this manuscript.

REFERENCES


Symes, W. W., 2008a, Approximate linearized inver-
sion by optimal scaling of prestack depth migration: Geophysics, 73, R23–R35.
Focusing the wavefield inside an unknown 1D medium: Beyond seismic interferometry

Filippo Broggiini\textsuperscript{1}, Roel Snieder\textsuperscript{1} & Kees Wapenaar\textsuperscript{2}
\textsuperscript{1} Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401, USA
\textsuperscript{2} Department of Geotechnology, Delft University of Technology, P.O. Box 5048, 2600 GA Delft, The Netherlands

ABSTRACT

With seismic interferometry one can retrieve the response to a virtual source inside an unknown medium, assuming there is a receiver at the position of the virtual source. In this paper, we demonstrate that for a 1D medium the requirement of having an actual receiver inside the medium can be circumvented, going beyond seismic interferometry. This is accomplished using inverse scattering theory. We show that the wavefield can be focused inside an unknown medium with independent variations in velocity and density using reflection data only.

Key words: acoustic, scattering, wave propagation, crosscorrelation, reflectivity

1 INTRODUCTION

There are different ways to reconstruct the wave field excited by a hypothetical source in the interior of an unknown medium. First, with seismic interferometry it is possible to retrieve the response to a virtual source inside the medium, assuming there is a receiver at the position where the virtual source is to be created and assuming the medium is surrounded by sources. The medium parameters need not be known. Second, in this paper we show that, with one-dimensional inverse scattering theory, the response to a virtual source inside the medium can be obtained from reflected waves recorded at one side of the medium. We demonstrate that the second approach is possible for the one-dimensional situation without knowing the medium parameters. This is fascinating because it allows one to obtain the same virtual source response as with seismic interferometry (including all scattering effects), but without the need of having a receiver at the virtual source location. An essential element of this approach is to use a complicated incident wave that is designed to collapse onto a point inside the medium at a specified time.

Seismic interferometry, inverse scattering, and focusing are subjects usually studied in different research areas such as seismology (Aki and Richards, 2002), quantum mechanics (Rodberg and Thaler, 1967), optics (Born and Wolf, 1999), non-destructive evaluation of material (Shull, 2002), and medical diagnostics (Epstein, 2003). Seismic interferometry (Wapenaar et al., 2005; Curtis et al., 2006; Schuster, 2009) is a technique that allows one to reconstruct the response between two receivers from the cross-correlation of the wavefields measured at these two receivers that are excited by uncorrelated sources surrounding the studied system. This technique is also known as the virtual source method (Bakulin and Calvert, 2006); this refers to the fact that the new response is reconstructed as if one receiver had recorded the response due to a virtual source located at the other receiver position. Inverse scattering (Chadan and Sabatier, 1989; Gladwell, 1993; Colton and Kress, 1998) is the problem of determining the perturbation of a medium (e.g., of a constant velocity medium) from the field scattered by this perturbation. In other words, one wants to reconstruct the properties of the perturbation from a set of measured data. Exact inverse scattering takes into account the nonlinearity of the inverse problem, but it also presents some drawbacks: it may be improperly posed from the point of view of numerical computations (Dorren et al., 1994), and it requires data recorded at locations usually not accessible due to practical limitations. In this paper, the term focusing (Rose, 2001, 2002b) refers to the technique of finding an incident wave that collapses to a spatial delta function $\delta(z - z_0)$ at the location $z_0$. 

\[ \frac{1}{\sqrt{\Delta t}} \]

\[ \frac{1}{\sqrt{\Delta t}} \]
2 WAVEFIELD FOCUSING

Figure 1 illustrates a numerical scattering experiment in a one-dimensional acoustic medium where an impulsive source is placed at the position $z = 2.44$ km in the model shown in Figure 2. Note that velocity and density vary independently in depth. The acoustic wave equation is $LG(z, z_0, t) = -\delta(z - z_0)\frac{d}{dt}\delta(t)$, with the differential operator $L \equiv \rho(z)\frac{d}{dz}(\rho(z)^{-1}\frac{d}{dz}) - c(z)^{-2}\frac{d^2}{dz^2}$. Here $z_0 = 2.44$ km and the initial condition is $G(z, z_0, t < 0) = 0$. The incident wavefield propagates toward the discontinuities in the model, interacts with them, and generates scattered waves. We use a time-space finite-difference code with absorbing boundary condition to simulate the propagation of the one-dimensional waves and to produce the numerical examples shown in this section. For a computational purpose, the source function $-\delta(z - z_0)\frac{d}{dt}\delta(t)$ is convolved with a band-limited time wavelet $S(t)$. The computed wavefield is shown in Figure 1 and it represents the causal Green’s function of the system, $G$. Causality ensures that the wavefield is non zero only in the region delimited by the first arrival (i.e., the direct waves). The slope of the lines, representing the first arrival, depends on the velocity model of Figure 2.

Due to practical limitations, we usually are not able to place a source inside the medium we want to probe, which raises the following question: Is it possible to create the wavefield illustrated in Figure 1 without having a real source at the position $z = 2.44$ km? An initial answer to such a question is given by seismic interferometry. This technique allows one to reconstruct the wavefield that propagates between a virtual source and other receivers located inside the medium (Wapenaar et al., 2005). We remind the reader that this technique yields a combination of the causal wavefield $G$ and its time-reversed version $G^\alpha$ (i.e., anti-causal). This is due to the fact that the reconstructed wavefield is propagating between a receiver and a virtual source. Conceptually speaking, without a real (physical) source, one must have non-zero incident waves on a receiver to create waves that emanate from that receiver. The fundamental steps to reconstruct the Green’s function are (Curtis et al., 2006)
is and it is consistent with the result of the scatter-

It does not require any knowledge of

shows a cross-section of the

Having knowl-

Lamb (i.e., the particle

2002a

reflected impulse response measured at

z

We next assume that we only have access to reflected

hence we cannot place any sources or receivers inside it.

certain portion of the medium we want to study and

same wavefield, but often we are not able to access a

function reconstruction technique is shown in Figure

4.

Causal part of the wavefield estimated by the

Green's function reconstruction technique when the receiver

Figure 4. Causal part of the wavefield estimated by the

Green's function reconstruction technique when the receiver

located at

z

sl

and

zsr

indicate the

two real sources. 

zvs

shows the virtual source location.

Figure 3. Diagram showing the locations of the real and

virtual sources for seismic interferometry. 


1. measure the wavefields

\[ G(z, sl, t) \]

and

\[ G(z, sr, t) \]

at a receiver located at

\( z \) (\( z \) varies from 0.5 km to 4 km) excited by impulsive

sources located at

\( z_{sl} \) and \( z_{sr} \) at both sides of the

perturbation (a total of two sources in 1D) as shown in Figure

3;

2. cross-correlate

\[ G(z, sl, t) \]

with

\[ G(z_{vs}, sl, t) \]

where

\( z_{vs} = 2.44 \) km and

\( vs \)

stands for virtual source;

3. cross-correlate

\[ G(z, sr, t) \]

with

\[ G(z_{vs}, sr, t) \];

4. sum the results computed at the two previous

points to obtain

\[ G(z, vs, t) + G^\prime(z, vs, t); \]

5. repeat steps 1 to 4 for a receiver located at a

different

\( z \).

The causal part of wavefield estimated by the Green's function reconstruction technique is shown in Figure

4 and it is consistent with the result of the scattering experiment produced with a real source located at

\( z = 2.44 \) km, shown in Figure 1.

We thus have two different ways to reconstruct the

same wavefield, but often we are not able to access a
certain portion of the medium we want to study and
hence we cannot place any sources or receivers inside it.
We next assume that we only have access to reflected
waves \( R(t) \) measured above of the perturbation, i.e. the
reflected impulse response measured at \( z = 0 \) km due
to an impulsive source placed at \( z = 0 \) km. This further
limitation raises another question: Can we reconstruct
the same wavefield shown in Figure 1 having knowledge only of the reflected waves \( R(t) \)? Since there are
no sources present inside the perturbation, we speculate
that the reconstructed wavefield consists of a causal
and an anti-causal part.

For this one-dimensional problem, the answer to
this question is given by Rose (2001, 2002a). He shows
that there exists a particular incident wave that
collapses the wavefield to a spatial delta function at the
desired location after it interacts with the medium, and
that this incident wave consists of a spatial delta function
followed by the solution of the Marchenko equation
(Chadan and Sabatier, 1989; Lamb, 1980).

The Marchenko integral equation is a fundamental
relation of one-dimensional inverse scattering theory. It is
an integral equation that relates the reflected waves
\( R(t) \) to the incident wavefield \( u(t, t_f) \) which will create
a focus in the interior of the medium and ultimately gives
the perturbation of the medium. The one-dimensional
form of this equation is

\[
0 = R(t + t_f) + u(t, t_f) + \int_{-\infty}^{t_f} R(t + t') u(t', t_f) dt',
\]

where \( t_f \) is the one-way travel time from \( z = 0 \) to the
focusing location. We numerically solve the Marchenko
equation and construct the particular incident wave that
focuses at \( t = 0 \) s at a location specified by \( t_f = 3 \)
s. Any appropriate numerical method to solve integral
equations can be used to compute \( u(t, t_f) \). Note that
solving Equation 1 does not require any knowledge of
the medium: all that is needed is the reflection response
\( R(t) \) and the one-way travel time \( t_f \). Next, we inject
the particular incident wave \( \delta(t + t_f) + u(-t, t_f) \), as
shown in Figure 5, at \( z = 0 \) km and compute the
time-space diagram shown in the top panel of Figure 6: it
shows the evolution in time of the wavefield when the
incident wave is injected at \( z = 0 \) into the model. We
emphasize that this method is data driven, hence the
true model is not needed to build the particular incident
wavefield. The time-space diagram shown in Figure 6 is
computed using the true model, but this is done only
to illustrate the physics of the focusing process. The
bottom panel of Figure 6 shows a cross-section of the
wavefield at time \( t = 0 \): the wavefield vanishes except
at location \( z = 2.44 \) km. Note, however, that the time
derivative of the wavefield in Figure 6 (i.e., the particle
velocity) is not focused at \( z = 2.44 \) km, hence the energy
is not focused at this location. For this particular model,
\( t_f = 3 \) s corresponds to spatial focusing at the same
location where we placed the virtual source in Figure 4.
In order to focus the wavefield at a prescribed location
\( z_f \) (and not at the prescribed one-way travel time \( t_f \)),
we need to know an estimate of the travel times of the first
arrivals, but no information about the density would be
required.

Focusing the wavefield  171
Figure 5. Particular incident wave at $z = 0$ km that focuses at $t = 0$ s at $z = 2.44$ km. It is composed by a delta function (at $t = -3$ s) plus the time-reversed solution of the Marchenko equation, $u(-t, t_f)$, where $t_f = 3$ s. This wave has an approximate duration of 3.5 s.

Figure 6. Top: At $z = 0$ km, we inject the particular incident wave in the model and compute the time-space diagram by forward modeling. We denote this wavefield as $K(z, t)$. Note that the waves continue to propagate after 3 s. Bottom: cross-section of the wavefield at $t = 0$ s.

Figure 6 does not yet resemble the wavefield shown in Figures 1 and 4. However, if we denote the wavefield in Figure 6 as $K(z, t)$ and its time-reversed version as $K(z, -t)$, we obtain the wavefield shown in Figure 7 by adding $K(z, t) + K(z, -t)$. With this process, we effectively go from one-sided to two-sided illumination because in Figure 7 waves are incident on the scatterer from both sides for $t < 0$ s. The incident waves are non zero for $-6 < t < -3$, but to facilitate a comparison with Figures 1 and 4 this time interval is not completely included in the figure. According to Figure 7, we create a focus at a location inside the perturbation without having a source or a receiver at such a location and without any knowledge of the medium properties; we only have access to the reflected impulse response measured above of the perturbation. With an appropriate choice of sources and receivers, this experiment can be done in practice, e.g., in an acoustic laboratory (Rose, 2002a). Burrige (1980) shows diagrams similar to Figures 6 and 7 and explains how to combine such diagrams using causality and symmetry properties. The upper cone in Figure 7 corresponds to the causal Green’s function $G$ and the lower cone represents the anti-causal Green’s function $G^a$. Note that a small amount of energy is outside the two cones. This is due to numerical inaccuracies in our solution of the Marchenko equation. The causal part of the trace at $z = 0$ km is the virtual source response that we obtained without using the model.

The anti-causal Green’s function $G^a$ follows from $G$ by time-reversal, hence it satisfies $LG^a = -\delta(z - z_0)\frac{d}{dt}\delta(t) = \delta(z - z_0)\frac{d}{dt}\delta(t)$, where we used that $L$ is invariant to time-reversal. Adding the differential equations for $G$ and $G^a$ shows that $G + G^a$ satisfies the homogeneous equation: $L(G + G^a) = -\delta(z - z_0)\frac{d}{dt}\delta(t) + \delta(z - z_0)\frac{d}{dt}\delta(t) = 0$. The term homogeneous means that the sum of $G$ and $G^a$ is source free. Hence, to focus the wavefield at the virtual source location, there must be a particular incident wavefield coming from another location. In fact, the knowledge that $G + G^a$ satisfies a homogeneous equation suggests that a combination of the causal and anti-causal Green’s functions is needed to focus the wavefield at a location where there is no real source (i.e., source free), as shown in Figure 7. Oristaglio (1989) shows a similar result, although he derives the difference (instead of the sum) of the causal and anti-causal Green’s functions due to a different definition of the Green’s functions.

The extension of the iterative process in two and three dimensions still needs to be investigated (Wapenaar et al., 2011). We speculate that focusing the wavefield at a prescribed location in two- and three-dimensional media requires an estimate of the primary traveltimes from the virtual source location. W give a first mathematical proof for a two-dimensional medium with density variations only.
3 CONCLUSIONS

There are three distinct ways to reconstruct the same physical wave state. A physical source, seismic interferometry, and inverse scattering theory allow one to create the same wave state (see Figures 1, 4, and 7) that focuses at a certain location $z_s$ (2.44 km in our examples). Seismic interferometry tells us how to build an estimate of the wavefield without knowing the medium properties, if we have a receiver at the same location $z_s$ of the real source in the scattering experiment of Figure 1 and sources surrounding the medium. Inverse scattering goes beyond this as it allows us to focus the wavefield inside the medium without knowing its properties, using only reflected waves recorded at one side of the medium. This can be done also when density and velocity vary independently. This fact is a new contribution because the inverse scattering theory of Rose (2001, 2002b) and others (Aktosun and Rose, 2002) did not deal with simultaneous changes in density and velocity, because one cannot retrieve two independent quantities from one time series of reflected waves. We also add another step, i.e., $K(z, t) + K(z, -t)$, which creates the response of the virtual source. This is an important insight that does not appear in any of the papers by Rose. We show that the interaction between causal and anti-causal wavefields is a key element to focus the wavefield where there is no real source.

We speculate that many of the insights gained in our one-dimensional framework are still valid in a two- and three-dimensional framework. An extension of this work to two or three dimensions would give us the theoretical tools for many useful practical applications. For example, if we knew how to create the three-dimensional version of the particular incident wavefield, we could focus the wavefield to a point in the subsurface to simulate a source at depth and to record data at the surface. This kind of application will be helpful for full waveform inversion (Brenders and Pratt, 2007) and subsalt imaging (Sava and Biondi, 2004), where waves that have traversed a strongly inhomogeneous overburden are of extreme importance.

ACKNOWLEDGMENTS

The authors would like to thank the members of the Center for Wave Phenomena, Kasper van Wijk, Andrew Curtis, and two anonymous reviewers for their constructive comments. This work was supported by the sponsors of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena.

REFERENCES

Oristaglio, M., 1989, An inverse scattering formula that uses all the data: Inverse Problems, 5, 1097–1105.
Creating a virtual source inside a medium from reflection data: a stationary-phase analysis

Kees Wapenaar¹, Filippo Broggini² & Roel Snieder²

¹ Department of Geotechnology, Delft University of Technology, P.O. Box 5048, 2600 GA Delft, The Netherlands
² Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401, USA

ABSTRACT
With seismic interferometry a virtual source can be created inside a medium, assuming a receiver is present at the position of the virtual source. Here we discuss a method that creates a virtual source inside a medium from reflection data, without needing a receiver inside the medium. Apart from the reflection data, an estimate of the direct arrivals is required. However, no information about the scatterers in the medium is needed. We analyze the proposed method for a simple configuration with stationary-phase arguments. We show that the retrieved virtual-source response correctly contains the multiple scattering coda of the inhomogeneous medium. The proposed method can serve as a basis for data-driven suppression of internal multiples in seismic imaging.

Key words: controlled source seismology, wave scattering and diffraction, interferometry, theoretical seismology.

1 INTRODUCTION
We discuss a new approach to creating the response to a virtual source inside a medium that goes beyond seismic interferometry. Broggini et al. (2011, 2012) propose the 1D version of this new approach. They show that, given the reflection response of a 1D layered medium, it is possible to obtain the response to a virtual source inside the medium, without the need to know the medium parameters. The method consists of an iterative scheme, akin to earlier work of Rose (2001). Interestingly, the response retrieved by this new method contains all scattering effects of the layered medium. Note that in order to obtain the same virtual-source response by seismic interferometry one would need a receiver at the position of the virtual source inside the medium, and real sources at the top and bottom of the medium. Hence, the advantage of the new approach over 1D seismic interferometry is that no receivers are needed inside the medium and that the medium needs to be illuminated from one side only. Broggini & Snieder (2012) speculate that the 1D method can be extended to three dimensions. This would imply that the 3D response to a virtual source in the subsurface could be retrieved from 3D reflection measurements at the surface, without knowing the parameters of the 3D medium. Hence, unlike for controlled-source interferometric methods (Schuster et al., 2004; Bakulin & Calvert, 2006), no receivers would be required in the subsurface, nor would the lack of sources illuminating the medium from below cause spurious multiples (Snieder et al., 2006).

Last year we made a first step towards generalizing the method of Broggini et al. (2011, 2012) to the 3D situation (Wapenaar et al., 2011). Using physical arguments we proposed an iterative scheme that indeed seems to transform the reflection response of a 3D medium into the response to a virtual source inside the medium. The proposed scheme requires, apart from the reflection data, an estimate of the direct arrivals between the virtual source and the acquisition surface. It is, in fact, through this traveltime curve that one specifies the location of the virtual source. Hence, the method is not fully model-independent. Note, however, that a model that relates direct arrivals to a source position can be much simpler than a model that explains all internal multiple scattering. In the proposed method the multiple-scattering part of the virtual-source response comes entirely from the reflection data.

The proposed method has not yet been proven mathematically (except for the 1D situation), nor have the limitations been exhaustively investigated. Here we
analyze the scheme for a simple 2D configuration with the method of stationary phase. We discuss the iterative procedure step-by-step and, en passant, repeat our physical arguments why the method is expected to converge to the virtual-source response, also in more complex situations.

2 THE CONFIGURATION

We consider a configuration of two parallel dipping reflectors in a lossless, constant velocity, variable density medium (Figure 1). The only reason for choosing a constant velocity is that all responses obey simple analytical expressions. However, the proposed scheme is not restricted to constant velocity media. We denote spatial coordinates as \( x = (x, z) \). The acquisition surface is located at \( z = 0 \) and is transparent (i.e., the upper half-space has the same medium parameters as the first layer). The first dipping reflector obeys the relation \( z = z_1 - ax \), with \( z_1 = 1000 \) m and \( a = 1/4 \). The red dot denotes the position of the virtual source, with coordinates \( x_{VS} = (x_{VS}, z_{VS}) = (100, 1400) \). The second reflector is chosen parallel to the first reflector, so that all mirror images of the source lie on a line perpendicular to the reflectors. This line obeys the relation \( z = z_1 + x/a \). The second reflector crosses this line at \( x = (150, 1600) \). The velocity in the medium is set to \( c = 2000 \) m/s. The density in the three layers are \( \rho_1 = \rho_2 = 1000 \) kg/m\(^3\), and \( \rho_3 = 5000 \) kg/m\(^3\). The reflection coefficients for downgoing waves at the two interfaces are \( r_1 = (\rho_2 - \rho_1)/\rho_2 + p_1 = 2/3 \) and \( r_2 = (\rho_3 - \rho_2)/\rho_3 + p_2 = -2/3 \), respectively. The reflection coefficients for upgoing waves are \(-r_1\) and \(-r_2\). The transmission coefficients for downgoing (+) and upgoing (−) waves are \( t_1^2 = 1 \pm r_1 \) and \( t_2^2 = 1 \pm r_2 \).

3 THE REFLECTION RESPONSE AND THE PRIMARY ARRIVALS

We introduce the Green’s function \( G(x, x_{VS}, t) \) as a solution of the wave equation \( L G = -\rho \hat{G}(x - x_{VS}) \frac{\partial^2 G}{\partial \omega^2} \), where \( L = \rho \nabla \cdot (\rho^{-1} \nabla) - c^{-2} \frac{\partial^2}{\partial \omega^2} \). Defined in this way, the Green’s function is the response to an impulsive point source of volume injection rate at \( x_{VS} \). The Green’s function is defined in the true medium. In the frequency domain, \( \hat{G}(x, x_{VS}, \omega) \) obeys the equation \( \hat{L} \hat{G} = -j\omega \rho \hat{G}(x - x_{VS}) \), with \( \hat{L} = \rho \nabla \cdot (\rho^{-1} \nabla) + c^2/\omega^2 \). Here \( j \) is the imaginary unit, \( \omega \) denotes angular frequency and the circumflex denotes the frequency domain. We write \( G = G^d + G^s \), where superscripts \( d \) and \( s \) stand for direct and scattered, respectively. We define the reflection response at the surface in terms of \( G^s \) via

\[
\hat{R}(x_R, x_S, \omega) \hat{\phi}(\omega) = \frac{2}{j\omega \rho_1} \frac{\partial G^s(x_R, x_S, \omega)}{\partial z_S} \hat{\phi}(\omega), \tag{1}
\]

for \( z_R = z_S = 0 \). Here \( \hat{\phi}(\omega) \) is the spectrum of the source wavelet \( s(t) \). Note (that apart from a factor \(-2\) \( \hat{R}(x_R, x_S, \omega) \hat{\phi}(\omega) \) represents the pressure at \( x_R \), related to a force source at \( x_S \), or, via reciprocity, the particle velocity at \( x_S \), related to a point source of volume injection rate at \( x_R \). Since we assumed that the acquisition surface is transparent, no surface-related multiples are present in the reflection response. Hence, \( \hat{R}(x_R, x_S, \omega) \hat{\phi}(\omega) \), or in the time domain \( R(x_R, x_S, t) \) \( = \hat{s}(t) \) (the asterisk denotes convolution), can be obtained from the measured reflection data by surface-related multiple elimination (Verschuur et al., 1992).

The aim of the proposed method is to use the deconvolved reflection response \( R(x_R, x_S, t) \) to retrieve the virtual-source response \( G(x, x_{VS}, t) \), with \( x_{VS} \) in the subsurface and \( x \) at the surface. As mentioned in the introduction, apart from the reflection response we need an estimate of the direct arrivals. For the constant velocity model of Figure 1a, the high-frequency approximation of the Fourier transform of the direct Green’s function \( G^d(x, x_{VS}, t) \) is given by \( \hat{G}^d(x, x_{VS}, \omega) = t_1^2 \rho_2 \hat{G}^d_0(x, x_{VS}, \omega) \), with

\[
\hat{G}^d_0(x, x_{VS}, \omega) = j\omega \exp\left\{ -j(\omega|\omega - \omega_{VS}|/c + \mu \pi/4) \right\}/\sqrt{8\pi|\omega||x - x_{VS}|/c}, \tag{2}
\]

with \( \mu = \text{sign}(\omega) \). Note that the scaled direct Green’s function \( \hat{G}^d_0(x, x_{VS}, \omega) \) requires only information on the propagation velocity \( c \) and not on the position and properties of the interfaces. Figure 1b shows \( \hat{G}^d_0(x, x_{VS}, \omega) \) and \( s_0(t) \), where \( s_0(t) \) is a zero-phase wavelet (here a Ricker wavelet with a central frequency of 20 Hz). In the following procedure the convolution with \( s_0(t) \) is optional, but in practice this wavelet will help to compensate for
the fact that the reflection response is deconvolved for the source wavelet \( s(t) \). If \( s_0(t) \) is included, it is essential that is zero phase.

4 INITIATING THE ITERATIVE PROCESS

The iterative process is started by defining an initial incident downgoing wave field at \( z = 0 \) as the time-reversed version of the direct arrivals, hence, \( p_0(x, t) = A^{(0)} G^0(x, x_{VS}, -t) * s_0(t) \), with \( A^{(0)} = 1 \), see Figure 2a for \( t < 0 \). The reflected upgoing wave field is obtained by convolving the incident field with the deconvolved reflection response and integrating over the source positions, according to

\[
p_0(x_R, t) = \int_{-\infty}^{\infty} [R(x_R, x, t) * p_0(x, t)]_{z=0} dx, \tag{3}
\]

for \( z_R = 0 \). We analyze this integral with stationary-phase arguments, starting with the response of the first reflector (a more detailed derivation is given in the Appendix). Figure 2b shows a number of rays, leaving different sources at the surface, reflecting at the first reflector, and arriving at one and the same receiver at \( x_R \). According to equation (3), these reflection responses are convolved with the initial incident field, of which the rays are also shown in Figure 2b (these are the rays that converge in \( x_{VS} \)). This convolution product is stationary for the source at \( x_{VS}^{(1)} \), where the rays of the incident field and of the reflection response have the same direction. Figure 2c shows a number of such stationary rays for different receiver positions. With simple geometrical arguments it follows that these rays cross each other at the mirror image of the virtual source with respect to the first reflector, i.e., at \( x_{VS}^{(1)} = (-100, 600) \). The travel-times of the convolution product are given by the length of the rays from \( x_{VS} \) to the surface, divided by the velocity. Hence, it is as if the response of the first reflector to the initial incident field originates from a source at \( x_{VS}^{(1)} \) (this is confirmed by the derivation in the Appendix).

This response is thus equal to \( A^{(1)} G^0(x_R, x_{VS}^{(1)}, t) * s_0(t) \), see Figure 2a (the first event for \( t > 0 \)). Here \( A^{(n)} \) is defined as the product of reflection and transmission coefficients contributing to the \( n \)th event of the reflection response \( R(x_R, x_S, t) \), hence \( A^{(1)} = r_1 \).

With similar arguments it follows that the response of the second reflector to the initial incident field apparently originates from a mirror image of the virtual source in the second reflector, i.e., at \( x_{VS}^{(2)} = (200, 1800) \). Hence, it is equal to \( A^{(2)} G^0(x_R, x_{VS}^{(2)}, t) * s_0(t) \), with \( A^{(2)} = t_1^- r_2 t_1^+ \), see Figure 2a (the second event for \( t > 0 \)). The multiple reflected responses to the initial incident field also apparently originate from mirror images of the virtual source, all located along the line \( z = z_1 + x/a \), see Figure 1a. For example, the first multiple is equal to \( A^{(3)} G^0(x_R, x_{VS}^{(3)}, t) * s_0(t) \), with \( A^{(3)} = -t_1^- r_2^2 r_1^+ t_1^- \) and \( x_{VS}^{(3)} = (500, 3000) \) (being the mirror image of \( x_{VS} \) in the mirror image of the first reflector with respect to the second reflector). Figure 2a (\( t > 0 \)) shows the total response \( p_0(x, t) \) to the initial incident field \( p_0(x, t) \). Because it has been derived with the method of stationary phase this response is free of artifacts. Of course in practice equation (3) needs to be evaluated numerically, which will introduce finite aperture artifacts, etc.
the first reflector (indicated by $r_1$) falls within the time window. Hence, for $k = 1$ we have $w(x, t)p_0^-(x, -t) = r_1 G_0(x, x_{VS}^{(1)}, -t) \ast s_0(t)$. Subtracting this from $p_0^+(x, t)$ gives, according to equation (4), the modified incident wave field $p_1^+(x, t)$. This is shown in Figure 3a ($t < 0$). Using equation (5) we evaluate the reflection response to this modified incident wave field. This reflection response is the superposition of $p_0^+(x, t)$, evaluated in the previous section, and the response to $-r_1 G_0(x, x_{VS}^{(1)}, -t) \ast s_0(t)$. Following the same reasoning as in the previous section, the first two terms of this additional response seem to originate from mirror images of $x_{VS}^{(1)}$ in the first and second reflector, hence, from $x_{VS}^{(1,1)} = x_{VS}$ (the original virtual source) and $x_{VS}^{(1,2)} = (400, 2600)$, respectively. The amplitude factors of these two terms are $-A^{(1)} r_1 = -r_1^2$ and $-A^{(2)} r_1 = -r_1 r_2 t_1^r r_1$, respectively. Higher order terms are evaluated in a similar way. Figure 3a ($t > 0$) shows the reflection response $p_1^-(x, t)$, with the amplitude factors of the different events indicated in the right margin. Note that within the time window (i.e., between the red curves) $p_1^-(x, t)$ is identical to $p_0^+(x, t)$, hence further iterations will not cause any changes. Already after one iteration we have achieved the desired situation mentioned at the beginning of this section. This is a consequence of analyzing the simple configuration of Figure 1a. For more complex configurations more iterations will be required.

6 CREATING THE VIRTUAL-SOURCE RESPONSE

After finalizing the iterative process, we define $p(x, t)$ as the superposition of the final incident and reflected wave fields. Because, for this example, iteration $k = 1$ was the final iteration, we have $p(x, t) = p_1^+(x, t) + p_1^-(x, t)$. This total field is shown in Figure 3a (all $t$). Note that $p(x, t)$ obeys the wave equation in the inhomogeneous medium of Figure 1a; Figure 3a is just the cross-section of this field at $z = 0$. Within the time window the field at $z = 0$ is antisymmetric in time (this was the design criterion for the iterative scheme). Hence, if we superpose the total field and its time-reversed version, i.e., $p(x, t) + p(x, -t)$, all events within the time window cancel each other, see Figure 3b. Note that because we consider a lossless medium, $p(x, t) + p(x, -t)$ obeys the wave equation. The causal part of this superposition is equal to $p_1^+(x, t) + p_1^-(x, -t)$ (Figure 3b, $t > 0$) and the acausal part is equal to $p_1^+(x, t) + p_1^-(x, -t)$ (Figure 3b, $t < 0$). Since time-reversal changes the propagation direction, it follows that the causal part is upward propagating at $z = 0$ and the acausal part is downward propagating at $z = 0$. The first arrival of the causal part of Figure 3b corresponds with the direct arrival of the response to the virtual source at $x_{VS}$ (Figure 1b). Given this last observation, combined with the fact that the causal part
is upward propagating at $z = 0$ and that the total field obeys the wave equation in the inhomogeneous medium and is symmetric, it is plausible that the total field is proportional to $G(x, x_{VS}, t) + G(x, x_{VS}, -t)$. This argumentation holds for more general situations, but we will check it here for the response in Figure 3b. From the procedure that led to this response, we find for the causal part (taking into account that $1 - r_1^2 = t_1^2 t_1^*$)

$$
 p_1^+(x, t) + p_1^+(x, -t) = t_1^2 t_1^* \left[ G_0^d(x, x_{VS}, t) + r_2 G_0^d(x, x_{VS}^{(2)} , t) - r_2 r_1 G_0^d(x, x_{VS}^{(3)} , t) \cdots \right] * s_0(t),
$$

with the virtual source position and its mirror images defined in sections 4 and 5. It is easily seen that for the configuration of Figure 1a, this is equal to $(t_1^2 / \rho_2) G(x, x_{VS}, t) * s_0(t)$. Hence, for the total field we have indeed

$$
 p(x, t) + p(x, -t) = t_1^2 / \rho_2 G_h(x, x_{VS}, t) * s_0(t),
$$

where $G_h(x, x_{VS}, t) = G(x, x_{VS}, t) + G(x, x_{VS}, -t)$. Note that because $G$ obeys the wave equation $LG = -\rho_2 \delta(x - x_{VS}/\beta(t))$ (section 3), $G_h$ obeys the homogeneous equation $LG_h = 0$. This is in agreement with the fact that $p(x, t) + p(x, -t)$ has been constructed without introducing a singularity at $x_{VS}$. $G_h$ is called the homogeneous Green’s function, after Porter (1970) and Oristaglio (1989) (but note that these authors take the difference instead of the sum of the causal and acausal Green’s functions because of a different definition of the source in the wave equation).

### 7 CONCLUDING REMARKS

For the configuration of Figure 1a we have shown that the iterative scheme of equations (4) and (5), combined with the process discussed in section 6, creates the response to a virtual source in an unknown medium from the reflection response at the surface and an estimate of the direct arrivals.

The iterative scheme of equations (4) and (5) is akin to an iterative solution of inverse scattering methods. As shown in equations (2.27), (3.16), and (5.18) of Burridge (1980), the Gel’fand-Levitan equation, the Marchenko equation and the Gopinath-Sondhi equations of inverse scattering in one dimension can be written (in symbolic notation) as

$$
 K - F + \int_W K R = 0,
$$

where $R$ denotes the reflection data, $F$ a known function that may depend on the reflection data, and $K$ the function that one solves for. The integral $\int_W$ denotes integration over a windowed time interval. The integral equation (8) can be solved by iteration by writing it as $K = F - \int_W K R$ and by inserting the left-hand side into the right-hand side (Ge, 1987). This gives the iterative solution

$$
 K_k = F - \int_W K_{k-1} R.
$$

The iterative system of equations (4) and (5) has the same structure as the iterative scheme (9). The connection between iteration, time-reversal and inverse scattering based on the Marchenko equation in 1D has been clarified in great detail by Rose (2001, 2002a,b). Brogini et al. (2011, 2012) show how to use the 1D scheme to create a virtual source in an unknown medium. Our present work aims at generalizing this to 2D and 3D media that are illuminated from above.

The stationary-phase analysis in this paper gives insight in the mechanism of the 2D iterative scheme and confirms the creation of the virtual-source response for a simple configuration. Following the physical arguments in section 6, it is plausible that the proposed methodology will also apply to more complex environments. Of course the proposed method will also have its limitations. The effects of a finite acquisition aperture, triplications, head waves, diving waves, fine-layering, errors in the direct arrivals, etc. need further investigation. A first numerical test by one of us (FB) with a variable-velocity syncline model shows promising results with respect to the handling of triplications. Errors in the estimated direct arrivals will cause defocusing and mispositioning of the virtual source (similar as in standard imaging algorithms), but we expect that these errors will not deteriorate the reconstruction of the internal multiples (which come from the response of the actual medium).

For those configurations for which the proposed methodology applies, the potential applications are fascinating. The method enables the retrieval of the response to a virtual source in an unknown medium, assuming an estimate of the direct arrivals is available. Since no actual receivers are needed inside the medium, virtual sources can be created anywhere. The virtual-source responses contain all internal multiples, hence the method could be used as a basis to image the subsurface without internal multiple ghosts (Wapenaar et al., 2012). Because the created virtual sources are independent of each other, the prediction and removal of internal multiples will not suffer from error propagation, unlike other imaging methods that aim at internal multiple suppression.

### REFERENCES


APPENDIX: DETAILS OF THE STATIONARY-PHASE ANALYSIS

We evaluate equation (3) with the method of stationary phase. In the frequency domain this equation reads

\[
\tilde{p}_0^+ (\mathbf{x}, \omega) = \int_{-\infty}^{\infty} \left[ \tilde{R}(\mathbf{x}, \mathbf{x}, \omega) \tilde{p}_0^- (\mathbf{x}, \omega) \right] e^{i \omega \mathbf{x} \cdot \mathbf{v}_S} d \mathbf{x},
\]

where the incident field \( \tilde{p}_0^- (\mathbf{x}, \omega) \) follows from complex conjugating the right-hand side of equation (2), according to

\[
\tilde{p}_0^- (\mathbf{x}, \omega) = -j \omega \frac{\exp \left\{ j (\omega |\mathbf{x} - \mathbf{x}_R^-| / c + \mu \pi / 4) \right\}}{\sqrt{8 \pi j |\mathbf{x} - \mathbf{x}_R^-| / c}} \delta_0 (\omega),
\]

with \( \mu = \text{sign} (\omega) \). For the reflection response we write

\[
\tilde{R}(\mathbf{x}, \mathbf{x}, \omega) = \sum_{n=1}^\infty \tilde{R}^{(n)}(\mathbf{x}_R, \mathbf{x}, \omega),
\]

where \( \tilde{R}^{(n)} \) represents the nth event of the reflection response. Using equation (1) we obtain

\[
\tilde{R}^{(n)}(\mathbf{x}, \mathbf{x}, \omega) = \frac{A^{(n)} z_R^{(n)} j \omega}{|\mathbf{x} - \mathbf{x}_R^{(n)}|} \frac{\exp \left\{ -j (\omega |\mathbf{x} - \mathbf{x}_R^{(n)}| / c + \mu \pi / 4) \right\}}{\sqrt{2 \pi j |\mathbf{x} - \mathbf{x}_R^{(n)}| / c}}.
\]

For \( \tilde{R}^{(1)} \), i.e. the primary response of the first reflector, \( A^{(1)} = r_1 \) and \( \mathbf{x}_R^{(1)} \) is the mirror image of \( \mathbf{x}_R \) with respect to the first reflector, see Figure A1. For \( \tilde{R}^{(2)} \), the primary response of the second reflector, \( A^{(2)} = r_1 r_2 t_1 \) and \( \mathbf{x}_R^{(2)} \) is the mirror image of \( \mathbf{x}_R \) in the second reflector. For \( \tilde{R}^{(3)} \), the first multiple, \( A^{(3)} = -t_1 r_2^2 r_1 t_1 \) and \( \mathbf{x}_R^{(3)} \) is the mirror image of \( \mathbf{x}_R \) in the mirror image of the first reflector with respect to the second reflector, etc. Equation (A.1) can be written as \( \tilde{p}_0^- (\mathbf{x}, \omega) = \sum_{n=1}^\infty \tilde{T}^{(n)} \), with

\[
\tilde{T}^{(n)} = \int_{-\infty}^{\infty} f(x) \exp \{ jk \phi(x) \} dx,
\]

where \( k = \omega / c \),

\[
f(x) = \frac{|\omega| A^{(n)} z_R^{(n)} \delta_0 (\omega)}{4 \pi |l_{VS}^{(n)}|^{1/2}}, \quad \phi(x) = l_{VS} - l_{R}^{(n)},
\]

with

\[
l_{VS} = |\mathbf{x} - \mathbf{x}_S| = \sqrt{\left( x - x_{VS} \right)^2 + \left( z_R^{(n)} \right)^2}.
\]

\[
l_{R}^{(n)} = |\mathbf{x} - \mathbf{x}_R^{(n)}| = \sqrt{\left( x - x_{R}^{(n)} \right)^2 + \left( z_R^{(n)} \right)^2}.
\]

The derivatives of the phase are

\[
\phi' (x) = \frac{x - x_{VS}}{l_{VS}} - \frac{x - x_{R}^{(n)}}{l_{R}^{(n)}},
\]

\[
\phi'' (x) = \frac{z_{VS}^2}{l_{VS}} - \frac{\left( z_R^{(n)} \right)^2}{l_{R}^{(n)}},
\]
Creating a virtual source from reflection data

The point $x_0^{(1)}$ depicted in Figure A1 obeys
\[ \frac{x_0^{(1)}(1) - x_{VS}}{z_{VS}} = \frac{x_0^{(1)}(1) - x_{R}^{(1)}}{z_{R}^{(1)}}. \]  
(A.10)

Generalized for $x_0^{(n)}$, this gives
\[ x_{0}^{(n)} = \frac{x_{VS}z_{0}^{(n)} - x_{R}^{(n)}z_{VS}}{z_{R}^{(n)} - z_{VS}}. \]  
(A.11)

Substitution into equation (A.8) gives $\phi'(x_0^{(n)}) = 0$, hence, $x_0^{(n)}$ is the stationary point of $\phi(x)$. According to equations (A.6) and (A.7) we have
\[ l_{VS}(x_0^{(n)}) = \frac{z_{VS}}{z_{R}^{(n)} - z_{VS}} |x_{R}^{(n)} - x_{VS}|, \]  
(A.12)
\[ l_{R}^{(n)}(x_0^{(n)}) = \frac{z_{R}^{(n)}}{z_{R}^{(n)} - z_{VS}} |x_{R}^{(n)} - x_{VS}|. \]  
(A.13)

In the main text we define $x_{VS}^{(n)}$ as a mirror image of $x_{VS}$, obtained in the same way as $x_{R}^{(n)}$ is obtained by mirroring $x_{R}$. This implies that $|x_{R}^{(n)} - x_{VS}| = |x_{R} - x_{VS}^{(n)}|$. This is illustrated for $n = 1$ in Figure A1. Hence, the stationary-phase approximation of equation (A.4) is (for large $|k|$)
\[ I^{(n)} \approx \sqrt{\frac{2\pi}{|k\phi'(x_0^{(n)})|}} f(x_0^{(n)}) \exp\{j(k\phi(x_0^{(n)}) + \mu\pi/4)\} \]  
\[ = j\omega A^{(n)} \exp\{-j(\omega|x_{R} - x_{VS}^{(n)}|/\epsilon + \mu\pi/4)\} \hat{s}_0(\omega), \]  
(A.14)

Hence,
\[ \tilde{p}_0(x_R, \omega) = \sum_{n=1}^{\infty} I^{(n)} = \sum_{n=1}^{\infty} A^{(n)} \hat{G}_0^{d}(x_R, x_{VS}^{(n)}, \omega) \hat{s}_0(\omega), \]  
(A.15)
with $\hat{G}_0^{d}(x_R, x_{VS}^{(n)}, \omega)$ as defined in equation (2), but with the source at $x_{VS}^{(n)}$. 

Figure 4. Configuration for the stationary-phase analysis of equation (10) for $n = 1$. 

The point $x_0^{(1)}$ depicted in Figure A1 obeys
\[ \frac{x_0^{(1)} - x_{VS}}{z_{VS}} = \frac{x_0^{(1)} - x_{R}^{(1)}}{z_{R}^{(1)}}. \]  
(A.10)

Generalized for $x_0^{(n)}$, this gives
\[ x_{0}^{(n)} = \frac{x_{VS}z_{0}^{(n)} - x_{R}^{(n)}z_{VS}}{z_{R}^{(n)} - z_{VS}}. \]  
(A.11)

Substitution into equation (A.8) gives $\phi'(x_0^{(n)}) = 0$, hence, $x_0^{(n)}$ is the stationary point of $\phi(x)$. According to equations (A.6) and (A.7) we have
\[ l_{VS}(x_0^{(n)}) = \frac{z_{VS}}{z_{R}^{(n)} - z_{VS}} |x_{R}^{(n)} - x_{VS}|, \]  
(A.12)
\[ l_{R}^{(n)}(x_0^{(n)}) = \frac{z_{R}^{(n)}}{z_{R}^{(n)} - z_{VS}} |x_{R}^{(n)} - x_{VS}|. \]  
(A.13)

In the main text we define $x_{VS}^{(n)}$ as a mirror image of $x_{VS}$, obtained in the same way as $x_{R}^{(n)}$ is obtained by mirroring $x_{R}$. This implies that $|x_{R}^{(n)} - x_{VS}| = |x_{R} - x_{VS}^{(n)}|$. This is illustrated for $n = 1$ in Figure A1. Hence, the stationary-phase approximation of equation (A.4) is (for large $|k|$)
\[ I^{(n)} \approx \sqrt{\frac{2\pi}{|k\phi'(x_0^{(n)})|}} f(x_0^{(n)}) \exp\{j(k\phi(x_0^{(n)}) + \mu\pi/4)\} \]  
\[ = j\omega A^{(n)} \exp\{-j(\omega|x_{R} - x_{VS}^{(n)}|/\epsilon + \mu\pi/4)\} \hat{s}_0(\omega), \]  
(A.14)

Hence,
\[ \tilde{p}_0(x_R, \omega) = \sum_{n=1}^{\infty} I^{(n)} = \sum_{n=1}^{\infty} A^{(n)} \hat{G}_0^{d}(x_R, x_{VS}^{(n)}, \omega) \hat{s}_0(\omega), \]  
(A.15)
Data-driven wavefield reconstruction and focusing for laterally-varying velocity models, spatially-extended virtual sources, and inaccurate direct arrivals

Filippo Broggini\textsuperscript{1}, Roel Snieder\textsuperscript{1} & Kees Wapenaar\textsuperscript{2}

\textsuperscript{1} Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401, USA
\textsuperscript{2} Department of Geotechnology, Delft University of Technology, P.O. Box 5048, 2600 GA Delft, The Netherlands

\textbf{ABSTRACT}
Seismic interferometry allows one to create a virtual source inside a medium, assuming a receiver is present at the position of the virtual source. We discuss a method that creates a virtual source inside a medium from reflection data measured at the surface, without needing a receiver inside the medium and, hence, presenting and advantage over seismic interferometry. In addition to the reflection data, an estimate of the direct arrivals is required. However, no information about the medium is needed. We analyze the proposed method for a simple configuration using physical arguments based on the stationary-phase method and show that the retrieved virtual-source response correctly contains the multiples due to the inhomogeneous medium. We illustrate the method with numerical examples in lossless acoustic media with laterally-varying velocity and density and take into consideration finite acquisition aperture and spatially-extended virtual sources. We examine the reconstructed wavefield when a macro model is used to estimate the direct arrivals.

\textbf{Key words:} focusing, wavefield reconstruction, stationary-phase, inverse scattering, finite aperture, virtual source, seismic interferometry, multiples

\section{INTRODUCTION}
We propose and discuss a new approach to reconstruct the response to a virtual source inside a medium. Controlled-source interferometric methods (Schuster, 2009; Bakulin and Calvert, 2006) allow us to retrieve such a response without the need to know the medium parameters, but these methods require a receiver at the location of the virtual source in the subsurface and assume the medium is surrounded by sources. Our new approach removes the constraint of having a receiver at the virtual source location and is based on an extension of the 1D theory proposed previously by Broggini et al. (2011, 2012); Broggini and Snieder (2012). They show that, given the reflection response of a 1D layered medium, it is possible to reconstruct the response to a virtual source inside the medium, without having a receiver at the virtual source location and without knowing the medium. The insights for this method were based on the work of Rose (2002). An essential element of this approach is to use a complicated incident wave that is designed to collapse onto a point inside the medium at a specified location and time.

Wapenaar et al. (2011a) made a first attempt to generalize to 3D media the method of Broggini et al. (2011, 2012) and Broggini and Snieder (2012). They used physical arguments to propose an iterative scheme that transforms the reflection response of a three-dimensional medium into the response to a virtual source inside the unknown medium. Apart from the reflection data measured at the surface, the proposed method also requires an estimate of the direct arrivals between the virtual source location and the acquisition surface. These arrivals are a key element of the method because they specify the location of the virtual source in the subsurface. For this reason, the proposed method is not fully model-independent. However, a model that relates the direct arrival to a virtual source position is simpler than a model that correctly handles the multiples.
In the proposed method, the reflection data contributes to the multiple-scattering part of the virtual-source response.

As in seismic interferometry (Wapenaar et al., 2005; Curtis et al., 2006; Schuster, 2009), our goal is to retrieve the response to a virtual source inside an unknown medium, removing the imprint of a complex subsurface. This is helpful in situations where waves have traversed a strongly inhomogeneous overburden (e.g., in subsalt imaging, Sava and Biondi, 2004). In this paper, we demonstrate that the requirement of having an actual receiver inside the medium can be circumvented, presenting an advantage over seismic interferometry.

In the next section, following the work of Wapenaar et al. (2012), we analyze the iterative scheme for a simple 2D configuration with three parallel dipping reflectors characterized by variable-density and constant-velocity. We use physical arguments based on the stationary-phase method to show that the method converges and allows for the reconstruction of the wavefield originating from the virtual source location. In the subsequent section, we present numerical examples in lossless acoustic media with variable velocity and density for more complex configurations: a layered medium with varying density and velocity and a model with a syncline reflector. We discuss the influence of errors in the estimate of the first arrivals on the reconstructed wavefield. Such errors arise if a macro model (a routine product of velocity analysis) is used to compute the first arriving waveforms when such data are not available with other approaches, e.g., check shots or microseismic events. We then examine how the finite acquisition aperture and the limitations of the modeling code affect the results. While seismic interferometry usually deals with virtual point sources, this method allows us also to reconstruct the response to a spatially-extended virtual source. This feature has a potential application in beam migration (Gray et al., 2009). Finally, we show that the proposed scheme implicitly reconstructs the incident field that focuses the wavefield in time and space at the virtual source location.

2 STATIONARY-PHASE ANALYSIS

We discuss the proposed iterative scheme for a simple two-dimensional configuration. We use a geometrical approach to the method of stationary phase to solve the Rayleigh-like integrals, which yield the reflected response to an arbitrary incident field. We explain each step of the iterative procedure and emphasize the physical arguments that support our expectations for the method to converge to the virtual-source response.

2.1 Configuration

We consider a configuration of three parallel dipping reflectors in a lossless, constant-velocity, variable-density medium (Figure 1). We choose a constant velocity medium because the response obeys simple analytical expressions. The proposed iterative scheme is, however, not restricted to constant velocity media as we show in the next sections. We denote spatial coordinates as \( x = (x, z) \). The acquisition surface is located at \( z = 0 \) m and is transparent (i.e., the upper half-space has the same medium parameters as the first layer). The first dipping reflector obeys the relation \( z = z_1 - ax \) with \( z_1 = 2000 \) m and \( a = 1/3 \). The black dot denotes the position of the virtual source, with coordinates \( x_{VS} = (x_{VS}, z_{VS}) = (475, 3425) \) m. The second and third reflector are parallel to the first one, so that all mirror images of the virtual source lie on a line perpendicular to the reflectors. This line obeys the relation \( z = z_1 + x/a \). The first, second, and third reflectors cross this line at \( x_1 = (z_1, 1) \), \( x_2 = (x_2, z_2) \), and \( x_3 = (x_3, z_3) \), respectively. The relative position of the reflectors is chosen such that the Euclidean distance between the pairs \( (x_1, x_2) \) and \( (x_2, x_3) \) is 1000 m. The velocity of the medium is constant and equal to \( c = 2000 \) m/s. The densities in the four layers are \( \rho_1 = \rho_3 = 1000 \) kg/m\(^3\), \( \rho_2 = 5000 \) kg/m\(^3\), and \( \rho_4 = 4000 \) kg/m\(^3\), respectively. The reflection coefficients for downgoing waves at the three interfaces are \( r_n = (\rho_{n+1} - \rho_n)/(\rho_{n+1} + \rho_n) \), where \( n = 1, 2, 3 \) denotes the layer. The reflection coefficients for upgoing waves are \( -r_n \). The transmission coefficients for downgoing \( (-) \) and upgoing \( (+) \) waves are \( t_n^\pm = 1 \pm r_n \). Since the velocity is constant in this particular configuration, the reflection and transmission coefficients hold not only for normal incidence but for all the angles of incidence. Note that the large contrast between the density of the layers causes strong multiple reflections.

2.2 Primary arrivals

We introduce the Green’s function \( G(x, x_S, t) \) as a solution to the wave equation \( LG = -\rho \delta (x - x_S) \partial \partial t \), with \( L = \rho \nabla \cdot (\rho^{-1} \nabla) - c^{-2} \partial^2 \partial t \). Defined in this way, the Green’s function is the response to an impulsive point source of volume injection rate at \( x_S \) (de Hoop, 1995). Using the Fourier convention \( \hat{F}(\omega) = \int_{-\infty}^{+\infty} f(t) \exp(-j\omega t) dt \), the frequency domain Green’s function \( \hat{G}(x, x_S, \omega) \) obeys the equation \( \hat{L} \hat{G} = -j\omega \rho \delta(x - x_S) \), with \( \hat{L} = \rho \nabla \cdot (\rho^{-1} \nabla) + \omega^2/c^2 \). Here \( j \) is the imaginary unit and \( \omega \) denotes the frequency domain. We write \( \hat{G} = \hat{G}^d + \hat{G}^t \), where superscripts \( d \) and \( t \) stand for direct and scattered waves, respectively. As mentioned in the introduction, we need an estimate of the direct arrivals. For the constant velocity model of Figure 1, the high-frequency approximation of the Fourier transform of the direct Green’s function \( \hat{G}^d(x, x_{VS}, t) \)
2.3 Reflection response

To retrieve the virtual-source response $G(x, x_{VS}, t)$, we need the reflection response at the surface $R(x_R, x_S, t)$ of $s(t)$ in addition to an estimate of the direct arrivals. We assume that the acquisition surface is transparent, hence the reflection response does not include any surface-related multiples. For this purpose, $R(x_R, x_S, t) * s(t)$ can be obtained from reflection data measured at $z = 0$ after surface-related multiple elimination (Verschuur et al., 1992). Following Wapenaar and Berkhout (1993) (equation 15a), the reflection response can be derived from a Rayleigh-type integral:

$$
\hat{p}^- (x_R, \omega) = \int_{-\infty}^{\infty} \frac{2}{j\omega \rho_1} \left[ \frac{\partial \hat{G}^s(x_R, x, \omega)}{\partial z} \hat{p}^+(x, \omega) \right] dz, \quad (2)
$$

where $\hat{p}^+$ and $\hat{p}^-$ are the Fourier transform of the downgoing and upgoing wavefields, respectively. Hence, in the frequency domain, we define the reflection response in terms of the scattered Green’s function $\hat{G}^s$ via

$$
\hat{R}(x_R, x_S, \omega) \delta(\omega) = \frac{2}{j\omega \rho_1} \frac{\partial \hat{G}^s(x_R, x_S, \omega)}{\partial x} \delta(\omega), \quad (3)
$$

for $z_k = z_S = 0$ and after multiplying both sides by the spectrum of the source wavelet $s(t)$. Note that, apart from a factor of $-2$, $\hat{R}(x_R, x_S, \omega) \delta(\omega)$ represents the pressure at $x_R$ due to a force source at $x_S$, or, via reciprocity, the particle velocity at $x_S$ due to a point source of volume injection rate $\frac{\partial h(t)}{\partial t} * s(t)$ at $x_R$ (Berkhout, 1982).

2.4 Initiating the iterative process

We define the initial incident downgoing wavefield at $z = 0$ as the time-reversed version of the direct arrivals at the recording surface excited by the virtual source in Figure 2. Hence, the initial wavefield is $p_0^d (x, t) = G^d(x, x_{VS}, -t) * s(t)$ and is shown in Figure 3 with the label 1. The subscript 0 of $p_0^d (x, t)$ indicates the initial wavefield (or the 0th iteration). In Figure 3, we also define two traveltime curves indicated by the solid black lines. The upper curve follows directly after the initial incident wavefield $p_0^d (x, t)$ and the lower curve is defined as the time-reversal of the upper curve (hence, it is placed just before the direct arrival of Figure 2). These two curves allow us to define a window function

$$
w(x, t) = 1 \quad \text{between the solid black lines of Figure 3},
$$

$$
w(x, t) = 0 \quad \text{elsewhere}. \quad (4)
$$

This window function is a key component of the iterative scheme.

The reflected upgoing wavefield $p_0^u (x, t)$ is obtained by convolving the downgoing incident wavefield $p_0^d (x, t)$ with the deconvolved reflection response and integrating
Figure 3. Initial incident wavefield \((t < 0)\) and its reflection response \((t > 0)\), both measured at \(z = 0\). Initial incident wavefield is the time-reversal of Figure 2. We show the reflection response only until 2.5 s. The solid black lines denote the time of the direct arrivals and its time-reversed counterpart. These lines are repeated in the subsequent figures.

over the source positions:

\[
p^{-}_0(x_R, t) = \int_{-\infty}^{\infty} \left[ R(x_R, x, t) \ast p^+_0(x, t) \right]_{z=0} \, dx, \tag{5}
\]

for \(z_R = 0\). Equation (5) is the time-domain version of the Rayleigh integral described by equation (2). We discuss and solve this integral with geometrical arguments based on the method of stationary phase (a detailed mathematical derivation is given by Wapenaar et al., 2012). Figure 4 shows a number of stationary rays for different receiver positions. These rays are said to be stationary because the rays of the incident field (converging in \(x_{VS}\)) and of the reflection response (for the first reflector) have the same direction (see the appendix in Wapenaar et al., 2010). With simple geometrical arguments, it follows that these rays cross each other at the mirror image of the virtual source with respect to the first reflector, i.e., at \(x_{VS}^{(1)}\). The traveltimes of the convolution product are given by the length of the rays from \(x_{VS}^{(1)}\) to the surface, divided by the velocity. Hence, it is as if the response of the first reflector to the initial incident field originates from a source at \(x_{VS}^{(1)}\). This response is thus equal to \(r_1 G^d(x_R, x_{VS}^{(1)}, t) \ast s(t)\) and is shown as the event with label 2 in Figure 3. Following similar stationary-phase arguments, the response of the second reflector to the initial downgoing field apparently originates from a mirror image of the virtual source with respect to the second reflector, i.e., at \(x_{VS}^{(2)}\). This response is equal to \(r_2 G^d(x_R, x_{VS}^{(2)}, t) \ast s(t)\) (see the event with label 3 in Figure 3). However, the multiple reflected responses to the initial incident field also apparently originate from mirror images of the virtual source, all located along the line \(z = z_1 + x/a\) (see Figure 1). We derive these responses with the method of stationary phase, hence they are free of artifacts. However, later in this manuscript, we show numerical examples and we discuss finite aperture artifacts, tapering effects, error in the estimate of the direct arrivals, etc.

Figure 4. Analysis of the response of the first and second reflectors to the initial incident wavefield \(p^+_0(x, t)\). (a) Stationary rays for different receivers. The response of the first reflector seems to originate from \(x_{VS}^{(1)}\). (b) Stationary rays for different receivers. The response of the second reflector seems to originate from \(x_{VS}^{(2)}\).

2.5 Iterative process

We now discuss an iterative scheme, which uses the \((k-1)\)th iteration of the reflection response \(p^\ast -_{k-1}(x, t)\) to create the \(k\)th iteration of the incident field \(p^+_k(x, t)\). The objective is to iteratively update the incident field in such a way that, within the upper and lower solid black lines shown in Figure 3, the field is anti-symmetric in time. The meaning of this criterion will be evident in the next section, where we show how to reconstruct \(G(x, x_{VS}, t)\). The method requires a combination of time reversal and windowing and the \(k\)th iteration of the incident field is defined by

\[
p^+_k(x, t) = p^+_0(x, t) - w(x, t)p^\ast -_{k-1}(x, -t), \quad \text{for } x \text{ at } z = 0, \tag{6}
\]

where the time window \(w(x, t)\) is defined by equation (4). The reflection response is then obtained using equation (5), which we rewrite here as

\[
p^\ast -_{k}(x_R, t) = \int_{-\infty}^{\infty} \left[ R(x_R, x, t) \ast p^+_k(x, t) \right]_{z=0} \, dx, \tag{7}
\]

for \(x\) and \(x_R\) at \(z = 0\). The first and second iteration of the incident and reflected fields are shown in Figures 5 and 6, respectively. The events of \(p^+_2(x, t)\) labeled 2, 3, 4 in Figure 6 correspond to the events of \(p^+_1(x, t)\) labeled 4, 5, 6 in Figure 5 (multiplied by \(-1\)). For this particular configuration, the \(k\)th iteration of the incident field (for \(k > 2\)) is similar to \(p^+_2(x, t)\) and is composed by four events, as shown in Figure 6 for \(t < 0\). The events labeled 1 and 4 remain unchanged in the iterative process. The other two events (labeled 2 and 3) correspond to \(-A^{(2)}_k G^d(x_R, x_{VS(2)}, -t) \ast s(t)\) and \(-A^{(3)}_k G^d(x_R, x_{VS(3)}, -t) \ast s(t)\), respectively. The coefficient \(A^{(2)}_k\) varies at each iteration and is equal to the
The numbers identify the first six events in the total field. We show the reflection response only until 2.5 s.

2.6 Wavefield reconstruction from the virtual source

After showing that the method converges to the desired result, we define \( p_k(x, t) \) as the superposition of the partial sum of the geometric series \( a + ax + ax^2 + ax^3 + ax^4 + \cdots + ax^k \) where \( a = r_1^3 t_2 t_3 t_4 \) and \( x = r_1^2 \). The sum of the series converges because \( x < 1 \) and yields

\[
\sum_{k=0}^{\infty} ax^k = \frac{a}{1-x} = r_2 t_2^{-1},
\]

where we used that \( 1 - r_2^2 = t_1 t_4^2 \). The coefficient \( A_k^{(3)} \) of the third event (labeled 3 in Figure 6) is equal to \(-r_1 A_k^{(2)}\) and, hence, it converges to \(-r_1 r_2 t_2^2 \). Figure 7 shows the thirtieth iteration and, within the time window \( w(x, t) \), the wavefield is antisymmetric in time. This is the result we predicted when we described the iterative method. Note that the antisymmetry was the design criterion for the iterative scheme. Wapenaar et al. (2012) show that, for their simple configuration with one dipping layer, convergence is reached after one iteration.

The numbers identify the first seven events in the total field. We show the reflection response only until 2.5 s.
as shown in Figure 8. Note that $p(x,t) + p(x,-t)$ also obeys the wave equation because we consider a lossless medium. The causal part of this superposition corresponds to $p^+(x,t) + p^+(x,-t)$ and the anti-causal part is equal to $p^+(x,t) + p^-(x,-t)$, as shown in Figure 8 for $t < 0$ and $t > 0$, respectively. From a physical point of view, time-reversal changes the propagation direction. Hence, it follows that the causal part propagates upward at $z = 0$ and the anti-causal part propagates downward at $z = 0$. The first event of the causal part of Figure 8 has the same arrival time as the direct arrival.

We speculate that $G(x,x_{VS},-t)$ and $G(x,x_{VS},t)$ are proportional to the anti-causal and causal parts of Figure 8, respectively.

This deduction does not depend on any particular feature of the configuration used in this analysis, hence it should hold for more general situations. We will check its validity for the response in Figure 8. Following the steps that led to the field shown in Figure 8, we find for the causal part that

$$p_{30}(x,t) + p_{30}^+(x,-t) = t_1^+ t_2^+ \left( G^d(x,x_{VS},t) + r_2 G^a(x,x_{VS}(t),t) - \sum_{m} (r_1 r_3 - r_2^2) G_d^a(x,x_{VS}(t),t) \ldots) \right) * s(t),$$

with the virtual source position and its mirror images shown in Figure 1. For the configuration of Figure 1, expression (9) is proportional to the wavefield $G(x,x_{VS},t) * s(t)$ originated from the virtual source and recorded at the surface (with $t_1^+ t_2^+$ as the coefficient of proportionality). The directly modeled response to the virtual source is shown in Figure 7 and matches the causal part of the field shown in Figure 8. This is also illustrated in Figure 10, where we superposed the traces $t_1^+ t_2^+ G(x_{R0},x_{VS},t) * s(t)$ and $p^-(x_{R0},t) + p^+(x_{R0},-t)$ where $x_{R0} = (0,0)$.

For the total wavefield, we obtain

$$p(x,t) + p(x,-t) = t_1^+ t_2^+ G_h(x,x_{VS},t) * s(t),$$

where $G_h(x,x_{VS},t) = G(x,x_{VS},t) + G(x,x_{VS},-t)$. Note that $G(x,x_{VS},-t)$ obeys the same wave equation as $G(x,x_{VS},t)$, i.e., $LG(x,x_{VS},-t) = -\rho_0 \delta (x-x_{VS}) \frac{\partial^2}{\partial t^2} = \rho_3 \delta (x-x_{VS}) \frac{\partial^2}{\partial t^2}$. Hence, $G_h$ obeys the homogeneous equation $LG_h = -\rho_3 \delta (x-x_{VS}) \frac{\partial^2}{\partial t^2} = 0$. This is in agreement with the fact that $p(x,t) + p(x,-t)$ has been constructed without introducing a singularity (i.e., a real source) at $x_{VS}$. $G_h$ is called the homogeneous Green’s function, after Porter (1970) and Oristaglio (1989) (but note that these authors take the difference instead of the sum of the causal and acausal Green’s functions because of a different definition of the source in the wave equation).

Finally, we introduce a visual way to show the convergence of the proposed method, compute the energy of $p_h(x,t) + p_k(x,-t)$ within the time window $w(x,t)$, and plot it versus the number of iterations:

$$\text{Energy}(k) = \sum_{t} \sum_{x} w(x,t) \left[ p_h(x,t) + p_k(x,-t) \right]^2.$$ (11)

The quantity is shown in Figure 11 and the energy inside the time window clearly converges to zero, as confirmed by Figure 8. Note that this procedure is expected to converge because in each iteration the reflected energy is smaller than the incident energy. We consider the proposed method as a correction scheme that minimizes the energy inside the time window $w(x,t)$.

3 NUMERICAL EXAMPLES

In this section, we show numerical examples for two-dimensional lossless acoustic media with variable velocity and density. We consider two different configurations: two parallel dipping reflectors and a syncline.
Figure 11. Energy of $p_k(x,t) + p_k(x,-t)$ inside the time window $w(x,t)$ versus number of iterations $k$. A logarithmic scale (base 10) is used for the Y-axis.

model. We require that the wavefield focuses at a specific location, and hence the proposed method cannot be totally independent of knowledge about the medium. The iterative scheme requires the reflection response of the medium measured at the surface, complemented with independent information about the primary arrivals originated from the focusing location, to focus the acoustic wavefield inside the medium. The primary arrival wavefront can be estimated or measured in various ways: by forward modeling using a macro model, directly from the data by the Common Focusing Point method (CFP) (Thorbecke, 1997) when the virtual source is located at an interface, from micro-seismic events (Artman et al., 2010), or from borehole check shots. We consider a transparent acquisition surface (i.e., the upper half-space has the same medium parameters as the first layer), so that the data do not include surface-related multiples. In Section 2, we used wiggle plots because the response obeys analytical expressions and hence these kinds of plots allowed us to clearly show the waveform of the events. Here, we use raster plots to display the wavefields because they are the standard visualization method used to show seismic images.

### 3.1 Two dipping reflectors

We first consider a configuration of two parallel dipping reflectors, whose velocity is shown in Figure 12a. The density profile (not shown) has a similar behavior as the velocity profile and the densities of the three layers are $\rho_1 = 1000$ kg/m$^3$, $\rho_2 = 5000$ kg/m$^3$, and $\rho_3 = 7000$ kg/m$^3$, from the top to bottom layer. The contrast between the densities is responsible for strong multiples. As in the previous section, we first consider a virtual point source and, with the next example, we discuss the iterative method for a spatially-extended source. To start the iterative scheme, we use a macro model to compute the direct arrivals originating from the virtual source indicated by the black dot in Figure 12b. Here, we use a smooth version of the true velocity model, as the macro model, as shown in Figure 12b. Smooth models are used in standard practice to test the validity of imaging techniques and are a product of velocity analysis. The initial incident field $p_i^0(x,t)$ is displayed in Figure 13a. Following the procedure explained in Section 2, we inject $p_i^0(x,t)$ into the actual medium to compute the initial response $p_0^0(x,t)$. This can be accomplished by convolving the incident field with the reflection responses measured at the surface (deconvolved for the source wavelet $s(t)$) and summing over the sources. Due to limited aperture, we apply a spatial tapering to the incident field to avoid edge-diffractions. Then we apply the window function $w(x,t)$ of equation 4 and construct the first iteration of the updated incident wavefield $p_i^1(x,t)$ (see Figure 13b). After computing $p_1(x,t)$, we build the field $p_1(x,t) + p_1(x,-t)$ to reconstruct the response originating from the virtual source location. The causal part of this field is shown in Figure 14a. Label (1) indicates an incomplete cancellation of the events inside the window $w(x,t)$. This is due to numerical limitations of the modeling code and not to a lack of convergence: it is caused by numerical dispersion and the influence of spatial-tapering on the initial incident...
field. Label (2) points to the imprint of finite acquisition aperture on the amplitude of the reconstructed events. For this particular configuration, we achieve the final result after one iteration because further iterations will not cause any changes in the incident field. The amplitude of the data shown in Figure 14 are clipped to 70% of the maximum amplitude and this enhances the features indicated by the labels (1) and (2). The comparison between the two panels of Figures 14 shows that it is possible to reconstruct the full response to a virtual source inside the medium, including all internal multiples, only from it first arrivals and the reflection response of the medium measured at the surface. This is fascinating because it allows one to obtain the same virtual source response as with seismic interferometry (including all scattering effects) with a starting smooth model only, but without the need for a receiver at the virtual source location.

This proposed scheme also implicitly constructs the incident field that collapses the wavefield onto the virtual source inside the medium and such a field corresponds to the last iteration of \( p_k^1(x,t) \). To show the physics of the focusing process, we simulate the propagation of the updated incident field \( p_k^1(x,t) \) inside the actual medium. Figure 15 displays six snapshots extracted from this simulation. Panel d clearly shows that the wavefield collapses to a focus at the location indicated by the black dot in Figure 12a; whereas the other events cancel in this panel.

3.2 Configuration with a syncline

In this section, we examine a configuration with a syncline reflector, whose velocity is shown in Figure 19a. The density profile (not shown) has a similar behavior and the densities of the three layers are \( \rho_1 = 6500 \text{ kg/m}^3 \), \( \rho_2 = 1000 \text{ kg/m}^3 \), and \( \rho_3 = 7000 \text{ kg/m}^3 \), from
Data-driven wavefield reconstruction and focusing

Figure 15. Time snapshots extracted from the propagating wavefield when the field of Figure 13b is used as a source. Panel d shows the wavefield focused at the location indicated by the black dot in Figure 12a. Time is increasing from panel a to f.

the top to bottom layer. The wavefield associated with this model is more complicated than the ones shown in Section 3.1; in fact the direct arriving wavefront contains a triplication. To start the iterative scheme, we compute the direct arrivals originating from the virtual source using the macro model of Figure 19b. This is a smooth version of the velocity model in Figure 19a. The initial incident field $p_0^+ (x, t)$ is displayed in Figure 20a. Note that, due to the smoothing, the triplications are almost not present in this field (i.e., the time-reversed version of the direct arrivals). As in Section 3.1, we compute the initial response $p_0^- (x, t)$ injecting $p_0^+ (x, t)$ into the actual medium, apply the window function $w(x, t)$, and construct the first iteration of the updated incident wavefield $p_1^+ (x, t)$ (see Figure 20b). This field is more complex then the field associated with the model considered in the previous section.

We form the field $p_1 (x, t) + p_1 (x, -t)$ to reconstruct the response originating from the virtual source location. The causal part of this field is shown in Figure 21a. Labels (1) and (2) indicate an incomplete cancellation of events. As in the previous section, this is due to numerical limitations of the modeling code: numerical dispersion and the influence of spatial-tapering on the initial incident field. Also, for this configuration, we achieve the final result after one iteration. As in the previous examples, the amplitude of the data shown in Figure 21 are clipped to 70% of the maximum amplitude and this enhances the features indicated by the labels (1) and (2). The comparison between the two panels of Figures 21 shows that it is possible to reconstruct the full response to a virtual source inside the medium, including all multiples, using the direct arrivals computed using a smooth model.

We simulate the propagation of the updated incident field $p_1^+ (x, t)$ inside the medium of Figure 19a. Figure 22 displays six snapshots extracted from this simulation. Panel d shows that the wavefield collapses to a focus at the location indicated by the black dot in Figure 19a. We note that the approximate direct arrivals (computed with the smooth model) used to start the iterative process causes an imperfect virtual source as indicated by the artifacts around the virtual source. We see that the complex wavefield propagating inside the syncline (panels a,b,c,e, and f) annihilates at the focusing time as shown in panel d.

Finally, we repeat similar steps for the same velocity model but we now use a different virtual source, i.e., a $\wedge$-shaped source (as in the previous section). The square inset (bottom right corner) in Figure 19a shows

Figure 16. (a) Initial incident wavefield $p_0^+ (x, t)$. (b) First iteration of the incident wavefield $p_1^+ (x, t)$. This initial field will focus the wavefield at the virtual source location in Figure 12a at $t = 0$, when the two planar and parallel line sources replace the virtual source indicated by the black dot.
the shape of the spatially-extended virtual source considered here. The ∧-shaped source replaces the black dot used previously. In contrast to the first example of this section, we compute the first arrivals using the true model instead of the smooth macro model. We do this to show that the quality and degree of focusing at the virtual source location depends on the first arrivals. Since the direct arrivals are now exact, Figure 25d shows that the wavefield focuses well on the target source. The proposed scheme produces a better result with fewer artifacts in this situation (compared to Figure 22d) because we have a correct estimate of the first arriving wavefront. Furthermore, the reconstructed response shown in Figure 24a looks almost identical to the directly-modeled response shown in Figure 24b.

4 CONCLUSIONS

We proposed a generalization to two dimensions of the model-independent wavefield reconstruction method of Broggi et al. (2011, 2012). Unlike the one-dimensional method, which uses the reflection response only, the proposed multi-dimensional extension requires, in addition to the reflection response, independent information about the first arrivals.

The proposed data driven procedure yields the response to a virtual source (Figures 14a, 17a, 21a, 24a) and reconstructs correct internal multiples, without needing a receiver at the virtual source location and without needing detailed knowledge of the medium. The method requires (1) the direct arriving wave front at the surface originated from a virtual source in the subsurface, and (2) the reflection impulse responses for all source and receiver positions at the surface. The direct arriving wave front can be obtained by modeling in a macro model, from microseismic events (Artman et al., 2010), from borehole check shots, or directly from the data by the CFP method (Berkhout, 1997) when the virtual source is located at an interface. In our numerical examples, we used a smooth version of the true model to compute the direct arrivals. The required reflection impulse responses are obtained from seismic reflection data after surface-related multiple elimination (Verschuur et al., 1992) and deconvolution for the source wavelet. Furthermore, because the proposed method is non-recursive, the reconstruction of internal
multiples will not suffer from error magnification, unlike other imaging methods that aim at internal multiple suppression (Berkhout and Verschuur, 2005; Luo et al., 2011).

For a simple configuration with planar dipping reflectors, the stationary-phase analysis gives insight into the mechanism of the two-dimensional iterative scheme and confirms that the method converges to the virtual-source response. Following the physical arguments in section 2, we produced numerical examples to test the method in various configurations and showed that the full response to the virtual source can be successfully reconstructed. We showed that the method is not limited to point sources, but it competently handles spatially-extended virtual sources. This feature permits one to create and steer oriented beams originating at depth under a complex overburden that generates strong multiples. This beamforming process could have a potential use in beam migration techniques (Gray et al., 2009).

Errors in the estimated first arrivals (due to a smooth macro model) cause defocusing and a mislocalization of the virtual source (as in standard imaging algorithms). Such errors, however, do not affect the handling of the internal multiples and do not deteriorate their reconstruction. This reconstruction is handled by the ac-
Figure 22. Time snapshots extracted from the propagating wavefield when the field of Figure 20b is used as a source. Panel d shows the wavefield focused at the location indicated by the black dot in Figure 19a. Time is increasing from panel a to f.

virtual medium through the reflection data measured at the surface (which includes all the information about the medium itself). Note that also the virtual source method (Bakulin and Calvert, 2006) does not optimally focus the wavefield at the virtual source location, but Wapenaar et al. (2011b) show that the focusing can be improved by applying multi-dimensional deconvolution.

ACKNOWLEDGMENTS

The authors thank the members of the Center for Wave Phenomena for their constructive comments. This work was supported by the sponsors of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena.

REFERENCES


Figure 23. (a) Initial incident wavefield $p_0^+(x,t)$. (b) First iteration of the incident wavefield $p_1^+(x,t)$. This initial field will focus the wavefield at the virtual source location in Figure 19a at $t = 0$, when the $\wedge$-shaped virtual source replaces the black dot.

Figure 24. (a) Causal part of the superposition of the total field and its time-reversed version, $p_1(x,t) + p_1(x,-t)$. Label (1) indicates an incomplete cancellation of the events inside the window $w(x,t)$. (b) Directly-modeled full response to the $\wedge$-shaped virtual source shown in Figure 19a.
Figure 25. Time snapshots extracted from the propagating wavefield when the field of Figure 23b is used as a source. Panel d shows the wavefield focused at the location indicated by the \( \wedge \)-shaped symbol in Figure 19a. Time is increasing from panel a to f.


——–, 2012, Focusing the wavefield inside an unknown 1D medium - Beyond seismic interferometry: submitted to Geophysics.


Oristaglio, M., 1989, An inverse scattering formula that uses all the data: Inverse Problems, 5, 1097–1105.


Newton-Marchenko-Rose Imaging: Image reconstruction based on inverse scattering

Jyoti Behura†, Kees Wapenaar‡ & Roel Snieder‡
†Center for Wave Phenomena, Colorado School of Mines, Golden, Colorado
‡Delft University of Technology, Netherlands

ABSTRACT
Using only surface reflection data and first-arrival information, we generate up- and down-going wavefields at every image point using the algorithm of Rose (2002b,a) and Wapenaar et al. (2011, 2012a). An imaging condition is applied to these up- and down-going wavefields directly to generate the image. Since the above algorithm is based on exact inverse scattering theory, the reconstructed wavefields are accurate and contain all multiply scattered energy in addition to the primary event. As corroborated by our synthetic examples, imaging of these multiply scattered energy helps illuminate the subsurface better. We also demonstrate that it is possible to perform illumination compensation using our imaging algorithm that results in improved imaging at large depths.

Introduction

Wapenaar et al. (2011) propose a methodology for reconstructing the 3D impulse response for any “virtual source” in the subsurface using surface reflection data and the direct arrivals from the “virtual source” to the receivers on the surface. Their proposal is the 3D extension of the 1D iterative algorithm of Rose (2002b,a) who shows that in layered media, it is possible to focus all the energy at a particular time (or depth if the velocity is known) by using a complicated source signature. In other words, the complicated source signature is the impulse response between the surface location and the depth at which the wave is focused.

It is imperative to briefly discuss the pioneering work of V. Marchenko, R.G. Newton, and J.H. Rose on inverse scattering theory (Prosser, 1969; Gopinath & Sondhi, 1971; Burridge, 1980; Bojarski, 1981; Newton, 1981) that is instrumental in the development of the methodology of Rose (2002b,a) and Wapenaar et al. (2011, 2012a) and the imaging algorithm presented here. In 1D scattering theory, Marchenko’s integral equation (Marchenko, 2011) determines the relation between the wavefield in the interior of a medium and the reflected impulse response. It was originally derived for spherically symmetric media (and therefore 1D) and used only reflected waves. Newton (1981, 1982) derived a similar relation that uses all scattered waves (reflected and transmitted) in 2D and 3D media. In 1D, this relation is given by

\[
\begin{align*}
\mathbf{u}^{+\text{sc}}(t, e, x) &= \sum_{e'=\pm 1} R(t + e' x, -e', e) \\
&+ \sum_{e'=\pm 1, \tau' = -\infty} \int_{-\infty}^{\infty} R(\tau + e' x, -e', e) \mathbf{u}^{+\text{sc}}(\tau, e', x) \, d\tau,
\end{align*}
\]

where \( t \) is time, \( e \) is the direction of wave propagation, \( x \) is the 1D space, \( u^{+\text{sc}} \) represents the scattered wavefield, and \( R \) is the impulse response function. The above relation is known as the Newton-Marchenko integral equation (Newton, 1981, 1982) that solves for the wavefield inside a medium using the far-field scattered data. Newton used the 3D version of equation (1) to solve the inverse scattering problem for the Schrödinger wave equation. Rose et al. (1983) generalized the Newton-Marchenko equation to the scalar wave equation and also derived the equation for attenuative and time-dependent media (Rose et al., 1986). Later, Budreck & Ross (1992) derived the relation for waves propagating in non-attenuative elastic media. A physical explanation of the above inverse scattering theory was provided by Rose (2002b,a) who showed that the ideas of focusing and time-reversal in fact result in the Newton-Marchenko equation.

The first step involved in solving for the scattering potential is to solve the Newton-Marchenko integral equation and find the wavefield everywhere inside the medium. Newton (1981, 1982) solved the inverse scattering problem for the Schrödinger wave equation by combining the Newton-Marchenko inte-
1 ALGORITHM

Any seismic imaging algorithm consists of two steps—wavefield reconstruction and imaging condition. For example, RTM is a two-way imaging technique that utilizes wavefields reconstructed in time by accurately implementing the wave equation \cite{Baysal1983, Whitmore1983, McMechan1983} in a smooth velocity model. Under this technique, the source- and receiver-wavefields are reconstructed by forward-propagating the source signature and back-propagating the receiver recordings using a numerical solution of the wave equation (commonly a finite-difference algorithm). Wavefield reconstruction in RTM is followed by the application of an imaging condition (commonly cross-correlation) to image the reflectors. In other words, we find the similarity between the wavefield that is incident on the reflector and the wavefield that is reflected. The similarity will be high at the reflector position and low elsewhere.

In NMRI, the up- and down-going wavefields are constructed at every location in space using the recipe of \cite{Wapenaar2011, Wapenaar2012}. An imaging condition is applied to these two wavefields to obtain the image. The pseudo-code for NMRI is given in Algorithm 1.

In NMRI, as the complicated incident wavefield focuses at the imaging point, a reflection is generated depending on whether there is actually a reflector at that point in space. In the presence of a reflector, the incident wavefield generates a reflected wave going in the direction opposite of the incident wave; while in the absence of a reflector, no reflected waves are generated at the image point. Therefore, only at a reflector location the incident-wavefield coincides with the reflected-wavefield which gives rise to a non-zero zero-lag cross-correlation (Figure 1a). Note that a background velocity model is necessary to compute the first arrivals at the surface from an impulse at the image point.

Since the Newton-Marchenko equation is based on exact inverse scattering, the reconstructed wavefields contain all multiply-scattered energy. Therefore, the wavefields that are incident on and reflected from the scatterers contain all the multiply-scattered energy in addition to the direct transmitted arrival. Note the presence of these multiples in Figures 1a and 1b. Any multiply-scattered wave that is incident on a scatterer will also have a corresponding scattered wave occurring at the same time. In Figure 1a the multiply-scattered incident- and reflected-waves occur at the same time, while in Figures 1b they do not coincide in time. Hence, in addition to the primary wavefield, all multiply-scattered energy will also be imaged accurately using NMRI. Other advantages of NMRI are discussed later.

Algorithm 1 Algorithm for NMRI. Superscript “s” represents time reversal and “∗” depicts convolution. \( R \) is the reflection response.

\[
\text{for any } x,y,z \text{ in image space do}
\]

\[
\begin{align*}
\text{Compute initial incident wavefield } u_0 \text{ from first-arrivals} & \\
\begin{align*}
u_{1,2}^1 & \leftarrow u_0 \\
u_{1,2}^2 & \leftarrow 0
\end{align*}
\end{align*}
\]

\[
\text{repeat}
\]

\[
\begin{align*}
\text{Mute } u_{1,2}^s \text{ beyond first arrival} & \\
\text{Update incident wavefield:} & \begin{align*}
u_1^1 & \leftarrow u_0^1 - u_0^s \\
u_1^2 & \leftarrow u_0^1 + u_0^s
\end{align*}
\end{align*}
\]

\[
\begin{align*}
\text{Updated scattered wavefield:} & \\
u_2^1 & \leftarrow u_2^1 \ast R \\
u_2^2 & \leftarrow u_2^2 \ast R
\end{align*}
\]

\[
\text{until } u_{1,2} \text{ converge}
\]

\[
\begin{align*}
\text{Compute } u_{up}, u_{dn} \text{ using the recipe of Wapenaar et al. (2011,2012)} & \\
\text{Apply imaging condition to } u_{up} \text{ and } u_{dn}
\end{align*}
\]

\[
\text{end for}
\]
Ray-tracing was used in computing the first breaks for the layercake model and for Lena; for the fold model, the first breaks were computed using a finite-difference wave-propagation code in a smoothed version (Figure 2b) of the true velocity model. An incorrect background velocity model will result in incorrect positioning of the reflectors. Although not analyzed here, it would be interesting to study how the NRMI image would differ from the RTM image for an incorrect velocity model.

Note that NRMI produced a satisfactory image (Figures 1d, 2c, and 4d) in each case. Some steeply-dipping events were not imaged accurately for the fold model and for Lena because of the lack of illumination.

2 NMRI IN ACTION

Here, we present three synthetic data results to demonstrate the performance and effectiveness of NRMI. The layer-cake (Figure 1) and Lena (Figure 3) models are constant velocity, variable density models while the fold model (Figure 2) has a variable velocity and constant density. The data acquisition is a fixed surface spread in each case where the sources and receivers are at \( z = 0 \) m. The source and receiver spacing is 10 m in all the three acquisitions. Time sampling is also the same (0.004 s) in each case. The direct arrivals were muted from the shot gathers. Besides this, no other processing was performed on the data; the data contain all orders of internal multiples.
of these features. Application of a low-cut filter to Lena (Figure 4e) shows that many small details have been imaged in detail. Even though the reflection data is complicated (Figure 3b) and contains all orders of internal multiples, NMRI imaged the primary as well as all the scattered events appropriately.

Figure 2. (a) The original velocity model of the fold system and (b) the smoothed version used for imaging with the corresponding NMRI image (c). A constant density of 1 gm/cm$^3$ was used in generating the reflection data.

3 ADVANTAGES OVER RTM

True amplitude/AVA: NRMI is based on exact inverse scattering theory and therefore the reconstructed wavefield in the interior of the medium is accurate irrespective of the velocity and density distributions in the subsurface. Hence, the NMRI image should be closer to the true reflectivity of the subsurface. Angle gathers for NMRI can be generated in the same way as in RTM. Amplitude variation with angle (AVA) analysis (on angle gathers) for NRMI, however, should be more reliable because the wavefields are accurate.

Multiples are imaged: As mentioned above, since the wavefields in NMRI are reconstructed accurately, the image should be better than existing imaging algorithms. Also, all orders of multiples are reconstructed and imaged accurately. Reflectors not illuminated by the direct arrival might be illuminated by internal multiples (Fleury, 2012); these reflectors would be visible on the NMRI image but not on the RTM image. Figure 4a is generated using only primaries while figure 4b results from imaging both primaries and internal multiples, with the difference between the two shown in figure 4c. Clearly, from figure 4c, internal multiples have contributed to the image at many locations. Note that the internal multiples do not result in any random or spurious events in the image. This is also corroborated by the virtual-source imaging of internal multiples of Wapenaar et al. (2011) and Wapenaar et al. (2012a) who note that the image corresponds to the reflectivity of the model. Imaging of multiples also renders multiple (both surface-related and inter-bed) prediction and suppression unnecessary. Although, none of the examples presented here contain surface-related multiples, the theory underlying NMRI imposes no limitations on the type of multiples.

Illumination compensation: Illumination compensation can be performed by manipulating the amplitude of the leading delta function in the incident wavefield such that each primary arrival at the image point has the same amplitude. By using the same unit delta function for all first arrivals, we ensure that the reflector is equally illuminated from all incidence angles and has the same illumination at all depths. However, if the Green’s function is used instead of a uniform delta function, the image at larger depths is poor because of insufficient illumination. For example, Figure 4a generated using the correct Green’s function (as the first arrival) can be interpreted as the best possible RTM image (generated from data containing only primary reflections). The above argument explains why the NMRI image (Figure 4b) is significantly better than the best possible RTM image (Figure 4a). Moreover, AVA would become more reliable as all the incident waves have the same amplitude. However, the effect of using a delta function instead of the Green’s function on focusing and on the amplitudes of multiply scattered waves needs more thorough analysis.

Targeted imaging: Note that for the computation of the image at any location in the image space, we need the first breaks at the surface for an impulse at the image point. Since the computation of the first breaks (using ray-tracing or finite-differences) can be done inde-
Figure 3. (a) The density model used in imaging Lena. (b) A sample shot gather. A constant velocity of 2000 m/s was used for modeling and imaging.

independently for each image point, it is possible to perform targeted imaging using NMRI. The targeted imaging of Lena’s left eye is shown in Figure 4f (Wapenaar et al. 2011) also show that targeted imaging of internal multiples is possible using virtual-source wavefields.

*Computationally cheap*: The most expensive component of any imaging algorithm is the computation of the wavefield in the interior of the medium. In RTM, the source- and receiver-wavefields are computed by solving the wave equation numerically which makes the method expensive. In NMRI, wavefield computation is done through an iterative process. For each image point, only a single run of ray-tracing is necessary to compute the first arrivals which are used for an initial guess of the incident wavefield. Thereafter, both the reflected and incident-wavefields are updated using the iterative procedure described in Algorithm 1 (in our tests two iterations were enough in most cases). The updated scattered field is obtained by convolving the updated incident wavefield with the reflected impulse response (recorded data) for each iteration. The low computational cost of ray-tracing and a set of convolutions can make NMRI significantly faster than RTM. A thorough quantitative analysis, however, is necessary to ascertain this.

*High frequencies*: The cost of RTM increases significantly with increasing frequency content because the extrapolation grid has to be more finely sampled. Wavefield computation using NMRI, on the other hand, has no such limitation because the frequency content of the incident wave and the impulse response are only limited by the temporal nyquist limit.

*Highly parallelizable*: Since the image at each location in the image space can be computed independently, the algorithm is highly parallelizable in the image space. For example, if there are 100 grid points in the image space, 100 processes could be run simultaneously on a cluster to obtain the image.

*Anisotropy*: Wavefield extrapolation in anisotropic media using numerical methods is expensive. In addition, depending on the dispersion relation used, the wavefield can contain shear-wave artifacts and incorrect P-wave amplitudes. In NMRI, however, if the first breaks are computed using ray-tracing, then imaging in anisotropic media becomes extremely cheap compared to RTM. Moreover, the wavefields in NRMI are accurate in amplitude even if the medium exhibits velocity anisotropy.

4 DISCUSSION AND CONCLUSIONS

The first arrivals at the surface from an impulse at any image point can be computed either using ray-tracing or solving the wave equation numerically using finite-differences. If the true phase and amplitude of the first arrival is desired, one can use dynamic ray-tracing or gaussian-beam modeling. However, to produce an image with illumination compensation, it is enough to do kinematic ray-tracing followed by a convolution with a delta function (Rose, 2002b,a). If the background velocity field results in multipathing, one must make sure that the incident wavefield contains all multiple arrivals; if not, the incident wave would not produce a quality focus at the image point.

Wapenaar et al. (2012) also demonstrate that the phase of the direct arrival is important in obtaining the correct amplitude and phase of the retrieved impulse re-
Figure 4. The NMRI image after one (a) and two (b) iterations and their difference (c). In (a) and (b), the Green’s function is used for the first arrival. (d) The NMRI image after two iterations with illumination compensation, i.e. the same magnitude of delta function is used for all first arrivals. (e) Low-cut filtered version of (d). (f) Targeted imaging of Lena's left eye.
response. To achieve this one can generate the first arrivals by numerically solving the wave-equation. In order to compensate for illumination, one must normalize each incident wave with its energy content to make sure that all incident waves have similar energy. As shown above, in the absence of normalization, deeper reflectors will not be illuminated adequately.

For a relatively uncomplicated subsurface, ray-tracing should suffice; the numerical solution of wave equation might be necessary for intricate first arrivals. Ray-tracing, however, has one significant advantage: it is substantially cheaper than solving the wave equation numerically (especially in anisotropic media). If minimizing computational cost is desired, ray-tracing should be the preferred method for computing the first arrivals.

Implementation of NMRI requires the construction of special shot-point gathers, which we call common surface-point gathers. Here, each bin location in the survey is a shot-point and every other bin is a receiver. Besides using conventional shot gathers in their construction, reciprocity can be used to further populate the common surface-point gathers. If no real traces exist for the shot and receiver combination, the trace can be interpolated or just assigned a null value. To reduce computational cost, an aperture can also be defined as done in many existing imaging algorithms.

Newton-Marchenko-Rose Imaging, which is based on exact inverse scattering theory, shows promise in imaging complicated subsurfaces. Besides primaries, it can be used for illumination compensation and can image both surface-related and internal multiples. This should make NMRI useful for imaging poorly illuminated areas, especially underneath salt bodies. In comparison to RTM, NMRI has other important advantages, such as, it is potentially computationally cheaper, can image arbitrarily anisotropic media, can be used for targeted imaging, and should generate accurate AVA response.

ACKNOWLEDGMENTS

The implementation of NMRI was done using [freeDDS]. Discussions with Farnoush Forghani and Filippo Broggni were very useful. Support for this work was provided by the Consortium Project on Seismic Inverse Methods for Complex Structures at CWP.

References


Fault surfaces and fault throws from 3D seismic images

Dave Hale
Center for Wave Phenomena, Colorado School of Mines, Golden CO 80401, USA

Figure 1. Roughly planar (a) and conical (b) fault surfaces and fault throws computed automatically from a 3D seismic image. Vertical and horizontal image slices are shown in the background. Vertical fault throws are measured in ms because the vertical axis of the image is time. Each quadrilateral intersects exactly one edge in the 4 ms by 25 m by 25 m image-sampling grid.

ABSTRACT
A new method for processing 3D seismic images yields images of fault likelihoods and corresponding fault strikes and dips. A second process automatically extracts from those images fault surfaces represented by meshes of quadrilaterals. A third process uses differences between seismic image sample values alongside those fault surfaces to automatically estimate fault throw vectors. While some of the faults found in one 3D seismic image have an unusual conical shape, displays of unfaulted images illustrate the fidelity of the estimated fault surfaces and fault throw vectors.

Key words: seismic geologic faults surfaces throws

1 INTRODUCTION
Fault surfaces like those displayed in Figure 1 are an important aspect of subsurface geology that we can derive from seismic images. Fault displacements, also shown in Figure 1, are important as well, as they enable correlation across faults of subsurface properties.

In the context of exploration geophysics, fault throw, relative displacement up or down the dip of a fault, is usually more significant than fault heave, displacement along the strike of a fault. Moreover, fault throw vectors are usually more perpendicular to geologic layers, and therefore easier to estimate, than are fault heave vectors.

As described by Luo and Hale (2012), we can use estimated fault throw vectors to undo faulting. Figure 2 displays multiple fault surfaces and corresponding fault throws computed for a 3D seismic image, before and after this unfaulting process. After unfaulting, seismic reflections are more continuous across faults, suggesting that estimated fault throws are generally consistent with true fault displacements.

Before this unfaulting, we must first compute images of faults, extract fault surfaces from those images, and estimate fault throws.
Figure 2. Fault surfaces and fault throws for a 3D seismic image before (a) and after (b) unfaulting.
1.1 Fault images

Several methods for highlighting faults, that is, for computing 3D images of faults from 3D seismic images, are commonly used today. Some compute a measure of the continuity of seismic reflections, such as semblance (Marfurt et al., 1998) or other forms of coherence (Marfurt et al., 1999). Others compute a measure of discontinuity, such as variance (Randen et al., 2001; Van Berkel and Pepper, 2011), entropy (Cohen et al., 2006), or gradient magnitude (Aqrawi and Boe, 2011). All of these methods are based on the observation that faults may exist where continuity in seismic reflections is low or, equivalently, where discontinuity is high.

However, in small regions within 3D seismic images, continuity may be low for reasons unrelated to faults. Stratigraphic features such as buried channels are well highlighted in seismic images by low continuity. Low continuity is also be caused by incoherent noise that is stronger than weak seismic reflections. Even when a fault is present, seismic events may appear to be highly continuous when fault throws are approximately equal to the dominant period (or wavelength) of those events. Event continuity alone is insufficient to distinguish faults.

For these reasons, Gersztenkorn and Marfurt (1999) noted that any measure of continuity or discontinuity must include some form of averaging within vertical windows that should be longer when detecting faults than when detecting stratigraphic features. In effect, these averaging windows smooth together small regions of low continuity that are vertically aligned along faults with significant vertical extent. More recently, Aqrawi and Boe (2011) noted that such vertical smoothing of image gradient magnitudes (computed via Sobel filters) is desirable when highlighting faults.

However, faults are seldom vertical. When averaging any seismic attribute used to highlight faults, we should vary the orientation of this averaging to coincide with the strikes and dips of those faults. Neff et al. (2000) and Cohen et al. (2006) do this in their computation of fault images, as they scan over a range of fault orientations for each sample in a 3D seismic image. The computational cost of such scans can be high when, for each 3D image sample and for each possible fault orientation, one must process many samples [over 1300, in the example of Cohen et al. (2006)] within some box-shaped neighborhood.

1.2 Fault surfaces

To extract fault surfaces like those shown in Figures 1 and 2 from 3D images of faults requires additional processing, which again has been performed in various ways.

For example, Pedersen et al. (2002, 2003) and Pedersen (2007, 2011) developed the method of ant tracking to merge together small regions of low continuity in 3D fault images into larger fault surfaces.

Gibson et al. (2005) propose a multistage method of constructing larger fault surfaces by merging smaller ones, beginning with small surfaces that correspond to “local discontinuities” in 3D seismic images. Different methods for growing large fault surfaces from small initial surfaces have also been proposed by Admasu et al. (2006) and Kadlec et al. (2008); Kadlec (2011). In such methods, seismic interpreters can specify seed points from which to begin growing fault surfaces.

In a more general context, Schultz et al. (2010) describe a direct method for extracting so-called crease surfaces from 3D images without seed points. In one example, they extract surfaces corresponding to ridges in a 3D image of fractional anisotropy, which is computed from 3D diffusion tensor magnetic resonance images (DT-MRI) of the human brain. Their method of extracting surfaces works well for 3D images with ridges that are well-defined and continuous.

1.3 Fault throws

Methods for computing fault surfaces lead naturally to the problem of estimating relative displacements of geologic layers alongside such surfaces. Solutions to this problem are not trivial, in part because of the sinusoidal character of seismic waveforms alongside faults, which can cause apparent horizontal alignment of seismic events across faults even when fault throws are significant. Another difficulty is that fault throws typically vary within the spatial extent of any fault surface. Nevertheless, several authors have described solutions to the problem of estimating fault throws.

For example, Aurhammer and Toni (2005) demonstrate the use of local crosscorrelations computed in rectangular windows and a genetic algorithm with geological and geometrical constraints to match horizons extracted from both sides of faults in 2D seismic images.

Liang et al. (2010) also used local crosscorrelations to estimate fault throws, while simultaneously scanning over fault dips to determine the locations and orientations of faults in 2D seismic images.

Admasu (2008) addressed the problem of estimating fault throws from 3D images through a Bayesian matching of seismic horizons extracted alongside faults in vertical 2D image slices, with the matching for one 2D slice used as a guide for the matching in adjacent slices. This method requires that faults surfaces are approximately orthogonal to the 2D image slices used to compute the fault throws.

In a 3D solution to the problem, Borgos et al. (2003) correlated seismic horizons across faults by clustering into classes local extrema in various attributes computed from 3D seismic images. Carrillat et al. (2004) and Skov et al. (2004) show examples of analyzing fault displacements computed using this method. In another
3D solution, Bates et al. (2009) demonstrated a “geo-model time differential analysis method” for computing fault throws after automatic horizon tracking.

### 1.4 This paper

This paper contributes to solutions of all three of the problems described above: (1) computing 3D fault images, (2) extracting fault surfaces, and (3) estimating fault throws. The sequence of solutions proposed here was used to compute the fault surfaces and throws displayed in Figures 1 and 2. Although each of these three solutions was designed in conjunction with the others in the sequence, aspects of any one of them could be adapted to enhance other methods summarized above.

I first compute 3D fault images of an attribute I call fault likelihood. Much like Cohen et al. (2006), I scan over multiple fault strikes and dips to maximize this semblance-based attribute. However, the computational cost of the algorithm I use to perform this scan is independent of the number of samples used in the averaging performed for each fault orientation. In other words, I improve computational efficiency by eliminating the factor (of 1300 or more) equal to the number of samples in the windows described by Cohen et al. (2006).

I then use the resulting 3D images of fault likelihoods, dips and strikes to extract fault surfaces using a method that is similar to that proposed by Schultz et al. (2010). The fault surfaces shown in Figures 1 and 2 are ridges in 3D images of fault likelihood, and are represented by meshes of quadrilaterals. I have made no attempt to fill any of the small holes apparent in these surfaces, although such a filling process would be easy to implement because every quadrilateral is linked to its neighbors. The fact that holes are small is due to the continuity of ridges in the 3D images of fault likelihood.

Finally, I compute fault throws from differences in values of samples extracted from 3D seismic images alongside fault surfaces. The algorithm I use to compute fault throws is derived from a classic dynamic programming solution (Sakoe and Chiba, 1978) to a problem in speech recognition. That solution today is often called dynamic time warping and is here extended to find a spatial warping that best aligns samples of 3D seismic images alongside faults, as illustrated in Figure 2.

### 2 FAULT IMAGES

Whereas seismic horizons appear in 3D seismic images as coherent events, a fault appears less prominently as a curviplanar surface on which seismic events are discontinuous, yet correlated, with some displacement, from one side of the fault to the other. Therefore, a useful first step in extracting fault surfaces and estimating fault throws is to first compute 3D fault images in which faults are most prominent.

#### 2.1 Semblance

The method I use for this first step is based on semblance (Taner and Koehler, 1969), and is therefore similar to methods proposed by Marfurt et al. (1998). Like Marfurt et al. (1999), I compute semblances from small numbers (3 in 2D, 9 in 3D) of adjacent seismic traces, after aligning those traces so that any coherent events are horizontal.

This process is illustrated for 2D seismic image shown in Figure 3a. Let $g[i_1, i_2]$ denote such an image, an array indexed by two integers: $i_1$, for time or depth, and $i_2$, for inline distance. To enhance the visibility of weaker features in this image, I applied the gain function $\text{sgn}(\cdot) \log(1 + |\cdot|)$ to every image sample.

Using structure tensors (e.g., van Vliet and Verbeek, 1995; Weickert, 1999), I first compute for the gained seismic image $g$ a corresponding image of local reflection slopes $p[i_1, i_2]$ displayed in color in Figure 3b. I then define and compute structured-oriented semblance as

$$s[i_1, i_2] = \frac{\text{sgn}(\cdot) \log(1 + |\cdot|)}{s_d[i_1, i_2]},$$

where $\langle \cdot \rangle$ denotes some sort of smoothing (discussed be-
low), and

\[
\hat{g}[t_i, i_x; j_x] = g(t_i + p[t_i, i_x] j_x, i_x + j_x),
\]

\[
s_n[t_i, i_x] = \left\{ \frac{1}{2M_x + 1} \sum_{j_x = -M_x}^{M_x} \hat{g}[t_i, i_x; j_x] \right\}^2,
\]

\[
s_d[t_i, i_x] = \frac{1}{2M_x + 1} \sum_{j_x = -M_x}^{M_x} \left\{ \hat{g}[t_i, i_x; j_x] \right\}^2.
\]  

(2)

Structure-oriented semblance is therefore simply the square of an average of slope-aligned sample values \( \hat{g} \) divided by an average of the squares of those same values. The number of traces in the local windows used to compute these averages is \( 2M_x + 1 \); I choose \( M_x = 1 \) so that only three traces must be aligned when computing semblance numerators \( s_n \) and denominators \( s_d \).

This definition of structure-oriented semblance is easily extended to 3D images \( g[t_i, i_x, i_y] \). After computing local inline slopes \( p_x \) and crossline slopes \( p_y \), I compute semblance numerators and denominators using local windows of \( 9 = 3 \times 3 \) slope-aligned traces.

### 2.2 Smoothing

The smoothing denoted by \( \langle \cdot \rangle \) in equation 1 is an essential part of the semblance computation for two reasons. First, without this smoothing, semblances are unstable where the denominators in equation 1 are nearly zero, that is, where slope-aligned values \( \hat{g} \) are nearly zero.

The second reason is that discontinuities in seismic images corresponding to faults are most significant for strong reflections that may be separated by multiple periods or wavelengths. Some sort of smoothing is necessary to link together these localized regions in which semblance numerators \( s_n \) are much smaller than semblance denominators \( s_d \).

It is for this second reason that Gersztenkorn and Marfurt (1999) recommend the use of longer vertical smoothing windows when highlighting structural features such as faults, and shorter windows when highlighting stratigraphic features such as channels. In proposing a different gradient-based measure of discontinuity, Aqrawi and Boe (2011) likewise use a vertical smoothing of that measure for the same reason.

Figure 4a shows structure-oriented semblances computed using a vertical two-sided exponential smoothing filter. This smoothing filter is efficient and trivial to implement. An implementation in the programming language C++ (or Java) for input array \( x \) and output array \( y \), both of length \( n \), is as follows:

```c
float b = 1.0f-a;
float y1 = y[0] = x[0];
for (int i=1; i<n-1; ++i)
    y[i] = y1 = a*y1+b*x[i];
y[n-1] = y1 = (a*y1+b*x[n-1])/(1.0f+a);
for (int i=n-2; i>0; --i)
    y[i] = y1 = a*y1+b*y[i];
```

Figure 4. Semblances \( s \) (a) and fault likelihoods \( f \) (b).

The extent of smoothing is controlled by the parameter \( a \). In the example shown in Figure 4, \( a = 0.93 \), which for low frequencies approximates a Gaussian filter with half-width \( \sigma = 20 \) samples. (In practice, I specify \( \sigma \) and compute the corresponding parameter \( a \).) As for a Gaussian filter, the impulse response for a two-sided exponential filter is infinitely long but decays smoothly to zero.

This vertical smoothing of semblance numerators and denominators accounts for the vertical extent of features with low semblance \( s \) apparent in Figure 4a. To accentuate these features I define an attribute fault likelihood \( f \) by

\[
f = 1 - s^8.
\]  

(3)

The choice of power 8 is somewhat arbitrary; it simply increases the contrast between samples with low and high fault likelihoods, as shown in Figure 4b.

Although features in semblance and fault likelihood images shown in Figure 4 have significant vertical extent, these features are not well aligned with faults, because the faults are not vertical. To improve the fault likelihood attribute \( f \), we must instead smooth along the faults. Our problem is that we have not yet determined the locations or orientations of the faults.

### 2.3 Scanning

This sort of problem is common in seismic data processing, for example, when we must perform normal moveout corrections without knowing the moveout velocities.
Figure 5. Fault likelihoods computed for two different fault dips $\theta$, one positive (a) and the other negative (b), in the scan used to estimate fault dips.

A common solution is to perform a scan for multiple velocities to find the velocities that maximize some (often semblance-based) measure of alignment. Here I scan over fault dips $\theta$ to find those dips that maximize fault likelihoods $f$.

Figure 5 illustrates the results of non-vertical smoothing for two different fault dips $\theta$ in this scan. These examples show that fault likelihoods tend to be largest when smoothing of semblance numerators and denominators is performed along the faults, which are not vertical.

To perform this non-vertical smoothing efficiently for each fault dip $\theta$, I (1) shear both semblance numerator and denominator images horizontally to make faults with that dip appear to be vertical, (2) apply the simple vertical smoothing filter described above, and (3) unshear the smoothed images before computing their ratio.

A fault that is vertical after horizontal shearing is shorter than it was before shearing. I therefore scale the half-width $\sigma$ of the vertical smoothing filter by $\cos \theta$, to compensate for this shortening.

Note that the cost of the scan over fault dips does not depend on the extent of smoothing, which is controlled by the parameter $a$ in the recursive smoothing filter. This recursive filter is largely responsible for reducing the computational cost of this scan, relative to those described by Neff et al. (2000) and Cohen et al. (2006).

This cost reduction is especially significant when scanning over both fault dips $\theta$ and strikes $\phi$ for 3D seismic images. In these scans smoothing of semblance numerators and denominators must be two-dimensional, within planes spanned by fault strike and dip vectors.

In scanning over fault strikes, for each strike angle $\phi$, I rotate the semblance numerator and denominator images, to align the fault strike direction with either of the horizontal image axes. I then smooth the rotated images once horizontally along the fault strike direction before scanning over fault dips $\theta$. The computational costs of rotation and horizontal smoothing for each fault strike $\phi$ are therefore negligible compared to the cost of the scans over fault dips $\theta$. The cost of an entire scan over fault strikes and dips for a 3D image is dominated by a sequence of scans over fault dips for multiple 2D images.

An alternative to the sequence of rotation, horizontal smoothing, shearing and vertical smoothing described above is to implement the smoothing filters $\langle \cdot \rangle$ with a fast Fourier transform (FFT). In my implementations, such FFT-based smoothing filters are simpler, but about three times slower, than the sequence described above.

Computational cost is also a factor in my choice of semblance, which requires smoothing of only numerator and denominator images $s_n$ and $s_d$. Alternatives such as the normalized correlation coefficient (Rodgers and Nicewander, 1988) or the eigen-structure-based coherence described by Gersztenkorn and Marfurt (1999) would require smoothing of more images for each fault strike and dip in the scan.

Sampling of fault strike and dip angles in the scan requires computation of angle sampling intervals and specification of lower and upper bounds. Because the units for axes of seismic images are often different — time versus distance — I measure angles in sample coordinates, so that an angle of forty-five degrees corresponds to a slope of one sample per sample.

Suitable sampling intervals that avoid undersampling are $\Delta \phi = \frac{1}{\sigma_\phi}$ and $\Delta \theta = \frac{1}{\sigma_\theta}$, both measured in radians, where $\sigma_\phi$ and $\sigma_\theta$ denote half-widths of the smoothing filters in the strike and dip directions, respectively.

When scanning to compute fault likelihoods for all examples shown in this paper, I chose $\sigma_\phi = 4$ samples and $\sigma_\theta = 20$ samples. These smoothing filter half-widths yield sampling intervals $\Delta \phi \approx 7.2$ degrees and $\Delta \theta \approx 1.4$ degrees. Minimum and maximum fault strikes were -90 and 90 degrees, and minimum and maximum fault dips (measured from vertical) were -15 and 15 degrees, so that the numbers of fault strikes and dips scanned were $N_\phi = 26$ and $N_\theta = 22$, respectively. The size of the product $N_\phi N_\theta = 572$ highlights the practical need to reduce the computational cost of computing fault likelihoods for each of the fault orientations in the scan over possible fault orientations.

Recall that fault dip angles are measured with re-
Fault surfaces and fault throws

211

Figure 6. Fault likelihoods computed by scanning over fault dips \( \theta \), before (a) and after (b) thinning.

spect to sample coordinates. Using the simple approximation that one millisecond of time corresponds approximately to one meter of depth, the true magnitudes of dips for many of the faults apparent in Figure 2 are roughly 45 degrees. Using the same approximation, the maximum fault dip scanned was roughly 60 degrees.

2.4 Fault likelihoods

The purpose of the scan over fault strikes and dips is to find, for each image sample, the angles \( \phi \) and \( \theta \) that maximize the fault likelihood \( f \). I begin with a fault likelihood image \( f = 0 \). Then, for each orientation \( (\phi, \theta) \) in the scan, where the fault likelihood \( f(\phi, \theta) \) exceeds the maximum likelihood stored in \( f \), I update \( f \) and also save the corresponding strike \( \phi \) and dip \( \theta \). When complete, the results of this scan are images of maximum fault likelihoods and corresponding fault strikes and dips.

Figure 6a shows fault likelihoods computed with a scan over 22 fault dips for the 2D seismic image. Ridges of fault likelihood in this fault image generally coincide with faults apparent in the seismic image. These ridges can be found by simply scanning each row of the fault image, preserving only local maxima, and setting fault likelihoods elsewhere to zero. In effect, this process thins the fault image, reducing the number of image samples at which a fault might be considered to exist.

Figure 6b shows ridges extracted from the fault image of Figure 6a, after discarding any ridges with fewer than \( 2\sigma_\theta = 40 \) adjacent samples. Parts of some ridges, especially those with lower fault likelihoods, may not coincide with faults. At this stage I do not suppress these parts, although one might easily suppress some of them by thresholding fault likelihoods.

Instead, I keep all ridges with length sufficient to reliably estimate fault throws. Then, faults can be assumed to exist at locations where fault likelihoods are high and fault throws are non-zero. Because, after thinning, so few samples are involved, such filtering of fault ridges can be performed interactively.

It is significant that the scanning process used to compute images of fault likelihood also yields images of fault strikes and dips for which fault likelihood is maximized. Those fault strike and dip angles are especially useful when extracting fault surfaces from ridges in 3D images of fault likelihoods.

3 Fault surfaces

One can easily imagine how to extract fault curves from 2D fault images like the one shown in Figure 6a. For example, we might simply link together samples with non-zero fault likelihood in the thinned fault image of Figure 6b. We could then use samples of the seismic image on the left and right sides of the extracted fault curves to estimate fault throws.

It is more difficult to construct fault surfaces from 3D fault images. One problem is how to best represent a fault surface, which need not be aligned with any axis of the sampling grid for the 3D seismic image. For example, the roughly conical fault displayed in Figure 1b cannot be projected onto a plane, and therefore cannot be represented by a single-valued function (such as distance) of coordinates within that plane.

Also, the resolution with which we sample fault surfaces will be important later, when we compute fault throws. In that computation, we must be able to efficiently traverse upward and downwards along fault curves of constant strike as we analyze seismic image samples alongside fault surfaces. We must also be able to efficiently traverse left and right along fault traces for which time (or depth) is constant.

For these reasons, I represent each fault surface with an unstructured mesh of quadrilaterals (quads) like those shown in Figure 1.

3.1 Extracting quads from fault images

My first step in constructing quad meshes is to extract a set of quads, not yet connected, from the 3D image of fault likelihoods. That 3D image is analogous to the 2D image of fault likelihoods shown in Figure 6a.

As shown in Figure 7, each quad in a fault surface intersects exactly one edge of the 3D sampling grid for the fault image. Each of the four nodes of a quad lies
within exactly one cell of that grid. The coordinates of a quad node within any such cell are averages of the coordinates of intersections of the fault surface and edges of the image sampling grid.

I find edge intersections and compute their locations using a method similar to that described by Schultz et al. (2010). I assume that fault surfaces are ridges in 3D images of fault likelihoods, analogous to the fault curves apparent in the 2D images shown in Figure 6.

These ridges intersect edges of cells in the 3D sampling grid, and can be found by considering all such edges, one at a time. Each edge is defined by two adjacent samples in the 3D image of fault likelihood. Let \( f_1 \) and \( f_2 \) denote fault likelihoods for these two samples. We may at this point choose a threshold \( f_{\text{min}} \) and assume that faults can exist only if both \( f_1 \geq f_{\text{min}} \) and \( f_2 \geq f_{\text{min}} \). I chose \( f_{\text{min}} = 0.5 \) when extracting the fault image scan yielded more consistent normal vectors.

Finally, like Schultz et al. (2010), I assume that ridges exist between two adjacent samples with vectors \( \mathbf{h}_1 \) and \( \mathbf{h}_2 \) that point in opposite directions, so that \( \mathbf{h}_1 \cdot \mathbf{h}_2 < 0 \). Letting \( \mathbf{x}_1 \) and \( \mathbf{x}_2 \) denote the spatial coordinates of two samples for which this condition is true, I compute the location \( \mathbf{x}_c \) where a ridge intersects the edge between those two samples by linear interpolation:

\[
\mathbf{x}_c = \frac{[\mathbf{h}_2^T (\mathbf{h}_2 - \mathbf{h}_1)] \mathbf{x}_1 - [\mathbf{h}_1^T (\mathbf{h}_2 - \mathbf{h}_1)] \mathbf{x}_2}{(\mathbf{h}_2 - \mathbf{h}_1)^T (\mathbf{h}_2 - \mathbf{h}_1)}. \tag{6}
\]

The spatial coordinates of a quad node located within any cell of the sampling grid are the average of coordinates of all quad-edge intersections for that cell. Spatial coordinates of the quad node are averages of the coordinates of all quad-edge intersections for that cell. This averaging enables representation of a fault surface with sub-voxel precision. Therefore, to find the locations of the quad nodes, we must first find the intersections of the fault surface and edges of the 3D sampling grid.

Following Schultz et al. (2010), let \( \mathbf{g} \) and \( \mathbf{H} \) denote the gradient vector and Hessian matrix for either of two adjacent samples in the fault likelihood image, like those shown in Figure 7. I compute each gradient (vector of 1st derivatives) \( \mathbf{g} \) and Hessian (matrix of 2nd derivatives) \( \mathbf{H} \) using simple centered finite-difference approximations to partial derivatives, after Gaussian smoothing (with radius \( \sigma = 1 \) sample) of the fault image to attenuate high frequencies for which those approximations are poor.

Now let \( \mathbf{H} = \lambda_u \mathbf{u} \mathbf{u}^T + \lambda_v \mathbf{v} \mathbf{v}^T + \lambda_w \mathbf{w} \mathbf{w}^T \) denote the eigen-decomposition of \( \mathbf{H} \), where the eigenvalues are ordered so that \( \lambda_u \geq \lambda_v \geq \lambda_w \). At locations of ridges, the smallest eigenvalue \( \lambda_u \) should be negative, and I assume that faults can exist only between two samples for which this condition is true.

If this condition is indeed true, then the eigenvectors \( \mathbf{w} \) for these two samples should be orthogonal to any ridge that may exist between them. Like Schultz et al. (2010), I then compute for each of these two samples a vector \( \mathbf{h} \) defined by

\[
\mathbf{h} = (1 - \lambda) \mathbf{w} \mathbf{w}^T \mathbf{g}, \tag{4}
\]

where

\[
\lambda = \begin{cases} 
0 & \text{if } \lambda_u - \lambda_w > \epsilon, \\
\frac{(1 - \frac{\lambda_u - \lambda_w}{\epsilon})^2}{\epsilon} & \text{otherwise},
\end{cases}
\tag{5}
\]

and \( \epsilon \) is a small fraction of the square of the typical fault image sample value. For fault likelihoods in the range \([0, 1]\), I use \( \epsilon = 0.01 \). The purpose of this parameter is to smooth the transition of the factor \( \lambda \) from zero to one where the eigenvalues \( \lambda_u \) and \( \lambda_w \) are nearly equal.

Recall that the scan used to compute the 3D fault likelihood image yields corresponding estimates of fault strike and dip angles for every image sample. I therefore make two significant modifications to the process of computing vectors \( \mathbf{h} \).

First, from the fault strike and dip I compute a fault normal vector \( \mathbf{n} \). I then assume that a fault can exist only between two samples for which \( |\mathbf{n}^T \mathbf{w}| > 1/2 \). This condition ensures some consistency between two different estimates of the normal vector; the angle between \( \mathbf{n} \) and \( \mathbf{w} \) must be less than 60 degrees. This upper bound on angle is rather large because the eigenvector \( \mathbf{w} \) of \( \mathbf{H} \) tends to be a poor estimate of the fault normal vector.

Therefore, in a second modification, if this condition is satisfied, I replace the eigenvector \( \mathbf{w} \) in equation (4) with the fault normal vector \( \mathbf{n} \) computed from the estimated fault strike and dip. I experimented with using the eigenvectors \( \mathbf{w} \) instead, as in Schultz et al. (2010), and found that the fault strikes and dips obtained during the fault image scan yielded more consistent normal vectors.
dinates \(x_c\) computed for all quad-edge intersections in that cell.

This interpolation and averaging yields quads that coincide (with sub-voxel precision) with ridges in the 3D image of fault likelihoods, and implies that the four nodes of a quad need not be coplanar. Also, while interpolating and averaging to compute the spatial coordinates of quad nodes, I interpolate and average the corresponding fault likelihoods, strikes and dips.

By analyzing all edges in the sampling grid for 3D images of fault likelihood, strike and dip, I extract quads that intersect edges where faults may exist. Because quads extracted in this way share nodes, they may appear to be parts of larger fault surfaces when displayed. However, at this point in the process of fault surface extraction, the quads are not yet linked together to form a surface mesh. I have only a collection of quads, what is sometimes called “quad soup.” In the example shown in Figure 2, this soup contained 436111 quads.

Before linking quads together to form meshes that represent fault surfaces, I perform one more test for consistency. Each of the four nodes referenced by any quad has associated estimates of fault strike and dip that define a fault normal vector \(n\). If we let \(x_a, x_b, x_c\) and \(x_d\) denote the spatial coordinates of these four quad nodes (in either clockwise or counterclockwise order), then the vector cross product \(n = (x_a - x_c) \times (x_b - x_d)\) should be (approximately, because the quad nodes need not be coplanar) normal to the quad. I keep only quads for which the angles between the vector \(n\) and each of the normal vectors \(n\) for its nodes are all less than 30 degrees. I remove from the soup any quad that fails this test. In the example shown in Figure 2, 337986 quads remained after this test.

### 3.2 Linking quads

The next step in extracting fault surfaces is to link quads together to form a mesh. Each quad in such a mesh may have up to four quad neighbors, where two quads are neighbors if they share an edge between two quad nodes. In the illustration in Figure 7 each quad has exactly two neighbors.

In the quad soup obtained using the process described above, it is possible that some edges between two quad nodes may be shared by more than two quads. Due to the multiple tests for consistency used in that process, this situation is rare; but where it occurs I choose to link none of the quads that share such an edge. This choice implies that two fault surfaces extracted in this way cannot intersect precisely, although they may be separated by only one grid sample.

Let us define the orientation of a quad such that its normal vector \(n\) points toward a viewer that sees the quad nodes \(a, b, c\) and \(d\) labelled in counter-clockwise order. From the opposite side, those same nodes would appear to be labelled in clockwise order. We can then define back and front sides of a quad such that the quad normal vector \(n\) points into the quad’s back side and points out from its front side.

At this stage in the process of linking quads, there is no guarantee that quad neighbors are oriented consistently. The normal vectors of a quad and its quad neighbors may point in opposite directions. However, any difference in the orientations of quad neighbors can be easily detected and accounted for when determining which quads share an edge between two of their nodes.

After all links between quads and their neighbors have been found, I apply three more filters to eliminate quads and links that are inconsistent with any geologically feasible model of fault surfaces. These filters work much like the consistency checks applied to quads in the quad soup described above.

The first filter unlinks any quads that are folded on top of one another. Specifically, I unlink two quads if the dihedral angle between them (computed from their normal vectors) is less than 90 degrees. This filter ensures that the orientation of a fault surface does not vary too rapidly from one quad to the next.

The second filter unlinks two quads if either of them (1) has no other neighbors or (2) has only one neighbor on the opposite side. Such quads tend to appear as fins or bridges between two nearby fault surfaces. In unlinking such quads, I assume that fault surfaces must nowhere be too skinny, with a width or height of only one quad.

After the first two filters remove geologically infeasible links between quad neighbors, the third filter simply removes any quads that have no neighbors. In applying this filter, I require that any fault surface must have a size of more than one quad.

When applied to the quads extracted in Figure 2, these three filters caused 15976 of the 337986 quads to be removed from the quad soup. All of the 322010 quads that remained were linked to at least two quad neighbors.

### 3.3 Constructing oriented fault surfaces

After extracting quads and linking them together, the final step in extracting fault surfaces is to find collections of quads that are linked either directly as neighbors or recursively as neighbors of neighbors. These collections form quad meshes that represent fault surfaces.

I assume that fault surfaces are orientable, that they have topologically distinct front and back sides, unlike the surfaces extracted from medical images by Schultz et al. (2010). In other words, I assume that fault normal vectors can be chosen consistently for every quad in the surface, so that the front side of every quad coincides with the front side of the surface.

This orientability assumption may be neither valid nor necessary, but is convenient here because of another assumption that I make when estimating fault throws as
described in the following section of this paper. I compute fault throws as a vector field of displacements from the back side to the front side of a fault surface, and vice-versa. In that process I assume that fault throws vary smoothly within each fault surface, and this assumption is most easily enforced where all of the quads in a fault surface are oriented consistently. Therefore, when collecting quads to form fault surfaces, I flip the orientations of quads as necessary to be consistent with their neighbors.

The collection process begins with a loop over all quads in any order. I first construct a new fault surface containing any quad. Then I add neighbors of this first quad to the surface, flipping their orientations as necessary to be consistent with that of the first quad. Then I recursively add quad neighbors, if not already in the surface, again flipping their orientations as necessary. The first fault surface is complete when there exist no unlinked quad neighbors that are not already part of that surface.

The collection process then returns to the loop over quads. When I find a quad that is not yet part of a fault surface, I again construct a new fault surface with that quad and recursively add quad neighbors to it in the same way as for the first fault surface. When the loop over all quads is complete, every quad belongs to exactly one fault surface.

I determine whether or not a fault surface is orientable during the recursive collection of quads. If, while examining neighbors of a quad, I find a neighbor that is already part of the surface and that has an inconsistent orientation, then the surface is not orientable. Otherwise, if all quad neighbors have consistent orientations, the surface is orientable.

In the extraction of surfaces shown in Figure 2, I found 1922 surfaces, and all of them were orientable. This figure displays only the 20 largest fault surfaces, those with at least 2000 quads.

In another example, I found that 0.03% of surfaces extracted were not orientable. I make such surfaces orientable by simply unlinking any quad neighbors that are already part of the surface and that have inconsistent orientations. In effect, this unlinking makes a surface orientable by cutting it in a rather arbitrary way that depends on the recursive order in which I add quads to the surface. Better methods for choosing the cut may be possible, and may be important where a large number of surfaces must be cut to make them orientable.

Because we typically view faults from the hanging wall side (above), and not from the footwall side (below), one final orientation of fault surfaces is useful for display. If necessary, I flip the orientations of all quads in each fault surface so that the average of the normal vectors for those quads points upward, not downward. After this final orientation, when viewing fault surfaces from the hanging wall side, we see the front side of the surface.

The filtering based on fault sizes used to obtain the 20 surfaces shown in Figure 2 is just one example of the sort of filtering that is possible after constructing fault surfaces. We could also filter these surfaces based on their average strikes or dips, (Pedersen et al., 2003; Pedersen, 2007, 2011), their fault likelihoods, or any combination of statistics derived from attributes computed for the quads that comprise the surfaces.

4 FAULT THROWS

Conceptually the problem of estimating fault throws from 3D seismic images is a simple one. We must correlate seismic reflections on one side of the fault with those on the other side, and compute vector displacements between corresponding reflections. In practice, this problem is difficult for several reasons.

4.1 Difficulties

First, our resolution of faults is limited by the resolution of seismic images, so that reflections on one side of a fault may extend somewhat into the other side. (See, for example, the image displayed in Figure 3a.) As suggested by Liang et al. (2010), our correlation of reflections across faults must in some way mimic the visual correlation of experienced seismic interpreters. That is, we must estimate fault throws from coherent seismic reflections extending well away from a fault, not only those immediately adjacent to it.

A second difficulty is that fault throws vary within a fault surface and within any local windows that we might use to correlate seismic reflections. To ease this difficulty, Aurnhammer and Tönness (2005) used several geologic and geometric constraints in a generic algorithm to estimate fault displacements.

Still, inconsistency remains in estimating a fault throw from a window of image samples in which that throw may vary significantly (L. Liang, personal communication, 2011). This second difficulty is exacerbated by the fact that such windows must be at least as long as the as the longest fault throw vector to be estimated.

A third difficulty is related to the fact that we can best estimate fault throws from strong seismic reflections with high signal-to-noise ratios, but these may be far apart, with weaker and noisier reflections in between. Constraints (e.g., Aurnhammer and Tönnes, 2005) are therefore needed to ensure continuity of fault throws estimated between strong reflections.

A fourth difficulty lies in constraining fault throws to vary smoothly in both dip and strike directions within a fault surface. Imposing this constraint is difficult partly because a fault surface typically cannot be projected onto a plane and then represented as a single-valued function (such as distance) of coordinates within that plane. This means that we cannot simply estimate...
Figure 8. Fault throws computed for a fault surface in which one part of the surface lies in front of another part. For such surfaces, we cannot compute fault throws from footwall and hanging-wall images extracted alongside the fault.

Figure 9. Fault throws are computed from differences between image sample values (squares) on the footwall side and values (circles) on the hanging wall side of a fault. These sample values are slightly offset from vertical quads, which intersect horizontal edges in the image-sampling grid. Fault strike is perpendicular to the plane of this figure.

fault throws from 3D seismic images by correlating one 2D image extracted from the footwall side of a fault surface with another 2D image extracted from the hanging-wall side.

An example is shown in Figure 8, where part of a fault surface lies in front of another part of that same surface. This situation occurs often in the 20 fault surfaces shown in Figure 2. Another example is the roughly conical fault displayed in Figure 1b. For such surfaces we cannot extract 2D images from the footwall and hanging-wall sides of a fault surface, so we cannot use 2D crosscorrelations of such images to estimate smoothly varying fault throws.

4.2 Dynamic warping

I address the difficulties summarized above with an extension of a classic method for estimating relative shifts between two acoustic signals in the problem of speech recognition (Sakoe and Chiba, 1978). This method is today widely known as dynamic time warping. For the problem of estimating fault throws, the most important aspect of this method is that it estimates a time-varying (dynamic) shift between two sampled functions of time, without any local windows. Another important aspect of this method is that the change in the shift with time can be easily constrained with no additional cost.

In a separate paper (Hale, 2012) I propose an extension of the dynamic time warping algorithm to the problem of dynamic image warping, in which we seek to constrain estimates of relative shifts between two images to vary smoothly in all directions. The extension is an alternating sequence of vertical (top-down, bottom-up) and lateral (left-right, right-left) smoothings of differences in image sample values, followed by the classic dynamic time warping algorithm. The vertical and lateral smoothings are non-linear, but simple and computationally efficient.

Unfortunately, as implied by Figure 8, we cannot reduce the problem of estimating fault throws to that of finding an optimal warping between footwall and hanging-wall images. We can, however, adapt the image warping solution to the problem of estimating fault throws, by performing a similar sequence of vertical and lateral smoothings within fault surfaces like those displayed in Figure 8.

Recall that each quad in this fault surface intersects exactly one edge in the sampling grid of a 3D seismic image, as shown in Figure 7. Some quads intersect vertical edges, but I assume that most quads, like those in Figure 7, intersect horizontal edges of the sampling grid. This assumption is valid even for fault dip angles greater than than 45 degrees, measured from vertical, because seismic images are typically sampled more finely in vertical directions than in horizontal directions.

As illustrated in Figure 9, let us refer to quads intersecting vertical edges as horizontal quads, and quads intersecting horizontal edges as vertical quads, even though quads are rarely exactly horizontal or vertical. I make this distinction because I use only the vertical quads in the dynamic warping process used to estimate fault throws.

My first step in estimating fault throws is to compute for each vertical quad two sequences of squared differences between image sample values collected from both sides of a fault surface. This first step is similar to computing squared differences of sample values in image warping, except that here I obtain sample values by following fault dip vectors up and down the fault surface, as illustrated in Figure 9.

Each intersection of a fault with a horizontal edge of the sampling grid corresponds to one vertical quad in a fault surface, like that near the center of the fault surface in Figure 9. For each such vertical quad I first find the value of the nearest sample (the filled square) located
at the same time (or depth), with some small lateral offset, in the footwall side of the fault. The purpose of this small offset is to compensate for limited resolution in 3D seismic images of faults. In Figure 9 the offset is two samples.

I then compute and store with this vertical quad the squared differences between the one footwall sample value (the filled square) and all of the values (the circles) on the hanging wall side of the fault. These squared differences form a sequence that is indexed by vertical lag.

I use that sequence and those computed for all other vertical quads in the same fault surface as inputs to the dynamic warping algorithm to estimate vertical components of throw vectors from the footwall side to the hanging wall side of the fault. I then estimate the horizontal components of throw vectors by once again following the fault dip vectors up or down the fault. I repeat this process to estimate fault throws from the hanging-wall side to the footwall side of the fault. By estimating throws in both directions, we can check pairs of throw vectors for consistency. If we follow the fault throw vector from a sample on the footwall side of the fault to a sample on the hanging-wall side, and then follow the throw vector found there back to the footwall side, we should return to the first sample at which we began.

A less rigorous test is to simply reject throw vectors on opposite sides of a fault where their vertical components have the same sign. In other words, we may assume that a fault exists only where throw vectors on opposite sides of the fault have vertical components with different signs.

This assumption is valid for all fault throws shown in this paper. The vertical components of all fault throws illustrated in figures are positive because the throws shown are those from the footwall side to the hanging-wall sides of normal faults.

In summary, the dynamic warping process described above computes fault throws that minimize the sum of squared differences between image sample values on two sides of the faults, while constraining the rate at which the throws may change within the fault surface.

4.3 Unfaulted images

A good test of the fidelity of estimated fault throws is to use them to undo faulting apparent in the seismic images from which they were derived. Luo and Hale (2012) describe this unfaulting process in detail. Here I use this process only to illustrate the accuracy of fault surfaces and fault throws computed using the methods proposed in this paper.

Figure 2 provides one example in which fault surfaces have roughly planar shapes. Faulting and unfaulting are most significant in the inline sections, because faults with large throws have strike vectors that point approximately in the crossline direction. Throws for these faults vary somewhat in the strike direction while generally increasing with depth.

Vertical exaggeration in these sections makes the faults appear to be more vertical than they really are; an approximate time-to-depth conversion (with 1 s equivalent to 1 km) indicates that the dips of most faults are about 45 degrees from vertical.

Visual comparison of the continuity of reflections before and after unfaulting suggests that estimated fault throw vectors are generally accurate. However, one location where estimated fault throw appears to have the wrong sign is at about 1.7 s and 4.5 km in the crossline section.

4.4 Conical faults

Figures 10 and 11 show faults extracted from the same seismic image for a shallower portion of the subsurface with more chaotic structure. Faults extracted from this portion have roughly conical shapes, in which the apex of each cone lies above its base. Figure 1b displays a close-up view of one of these conical faults.

These conical shapes became apparent to me only after extracting fault surfaces, partly because I had never seen faults with such shapes before, and so did not recognize their appearance in horizontal and vertical slices of the 3D seismic image.

This experience highlights an important benefit in using an automated process to extract information from 3D seismic images. The process used here could not exclude such shapes simply because they were unexpected.

After recognizing the conical shapes of these faults, they are easily seen in 3D seismic images, even without the fault throws shown in Figures 10 and 11. In vertical sections, these faults appear to be hyperbolic, because a vertical slice (conic section) of a cone is a hyperbola. When interactively moving a vertical slice through the 3D image, one can clearly see these hyperbolas rising and falling as the image slice moves through the cones.

Moreover, even in this relatively chaotic part of the seismic image, reflections are more continuous in the unfaulted image than in the original image. Some discontinuities that remain may be due to my including only the largest fault surfaces when computing fault throws. Of the 4608 surfaces extracted, I computed throws for only the 30 largest surfaces, those with at least 2000 quads.

5 CONCLUSION

I developed the methods proposed in this paper as parts of a three-step process to (1) compute fault images of likelihood, strike and dip, (2) extract fault surfaces, and (3) estimate fault throws. I once hoped to skip the first two steps, to simply compute fault throws everywhere
Figure 10. Fault surfaces and fault throws for a 3D seismic image before (a) and after (b) unfaulting. The shape for many of these faults is roughly conical, and the two vertical sections intersect near the center of one of these conical faults.
Figure 11. Fault surfaces and fault throws for a 3D seismic image before (a) and after (b) unfaulting. Conical faults appear as hyperbolas in vertical seismic sections.
and then let faults be defined as locations where fault throws are significant. However I was unable to find a computationally feasible implementation of this potentially simpler one-step process.

It is significant that the scan in the first step yields images of fault strikes and dips for which fault likelihood is maximized. These estimates of fault orientations are useful in several consistency tests performed in the second step used to extract fault surfaces.

I chose the quad-mesh representation for those fault surfaces in part to facilitate the third step of estimating fault throws. Because throw vectors connect samples on one side of a fault to those on the other side, it is especially convenient that quads in the fault surface lie between two adjacent samples of the seismic image at the same time or depth. In addition, the quad mesh also provides up-down and left-right connectivity needed to implement the dynamic warping algorithm used to estimate vertical throws.

Most of the computation time in this three-step process lies in the first step, which currently requires a scan over all possible fault orientations. I improve the computational efficiency of this scan by using fast recursive smoothing filters within each potential fault plane, but further improvements may be worthwhile. My current implementation of this scan for about 500 fault orientations requires about two hours to process a 3D image of 1000^3 samples on a 12-core workstation.

Perhaps the biggest current limitation in this process is in its handling of intersecting faults. As discussed above, to simplify the extraction of fault surfaces and estimation of fault throws, I have incorrectly assumed that faults do not intersect. Further work is required to extend these two steps to properly account for fault intersections.

ACKNOWLEDGMENTS

In much of the research described in this paper I benefited greatly from discussions with others, including Luming Liang, Marko Maučec, Bob Howard, Dean Witte, and Anastasia Mironova. The 3D seismic image used in this study was graciously provided by dGB Earth Sciences B.V. through OpendTect.

REFERENCES


Unfaulting and unfolding 3D seismic images

Simon Luo & Dave Hale
Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401, USA

ABSTRACT
One limitation of automatic interpretation methods such as seismic image flattening is their inability to handle geologic faults. To address this limitation, we propose to combine a method for automatic image unfaulting with seismic image flattening. First, using fault surfaces and fault throw vectors estimated from an image, we interpolate throw vectors to produce a throw vector field, which we use to unfault the image. Then, we flatten the unfaulted image to obtain a new image in which sedimentary layering is horizontal and also aligned across faults. From this flattened unfaulted image, we can automatically extract geologic horizons.

Key words: seismic image unfaulting flattening interpretation interpolation

1 INTRODUCTION
Extracting isochronal geologic surfaces—geologic horizons of the same age—is a common problem in geophysics and geology. Such horizons are useful for interpretation of stratigraphic features and analysis of structural deformation, as well as interpolation and correlation of subsurface properties. A geologic horizon is assumed to have been initially deposited as a horizontal layer and subsequently subjected to faulting and folding. So in order to extract such a horizon, it is necessary to quantify faulting and folding apparent in images such as the one shown in Figure 1a.

Perhaps the most straightforward way to extract a geologic horizon is by manual picking. Manual picking is often used in conjunction with autotracking methods (e.g., Howard, 1990), which track seismic events by following local extrema or zero crossings in amplitude in a seismic image. Most autotracking methods are not fully automatic, and an experienced interpreter is often required, for example, to identify seed points of coherent events to track, or to correlate events across faults. An obvious disadvantage of manual interpretation is that the process can be slow, because human interaction is required. The advantage, however, is that an experienced interpreter can pick horizons in areas in which a fully automatic method might have difficulty. Such areas could arise from a combination of geological (e.g., faults, unconformities, and complex stratigraphy) and geophysical (e.g., imaging and processing artifacts, noise, and multiples) complications (R. Howard, personal communication).

Alternatives to horizon autotracking methods include volume interpretation methods (e.g., Stark, 1996), which, rather than tracking single events, process simultaneously an entire seismic volume. Automatic seismic image flattening (Stark, 2004; Lomask et al., 2006; Parks, 2010) is an example of a volume interpreta-
tion method. Automatic seismic image flattening could potentially identify all horizons in an image, but the method is unable to match horizons across faults unless additional information (e.g., fault throw) is provided. Moreover, most automatic flattening methods are limited to only vertical shearing of an image, but images with non-vertical faults such as the one shown in Figure 1a clearly cannot be flattened by vertical shear alone.

A fully automatic method (e.g., Tnacheri and Bearnth, 2007) for extracting geologic horizons is ideal. Toward this end, we propose an automatic method that can be used to extract all geologic horizons in an image, consisting of two steps: image unfaulting followed by image unfolding (i.e., image flattening). To unfault an image, we first use the method described by Hale (2012) to estimate fault locations and fault throw vectors, displacement vectors along the dip direction of a fault surface. Then, using estimated fault throw vectors, we unfault the image. For example, we used fault throw vectors estimated from the seismic image shown in Figure 1a to obtain the unfaulted image shown in Figure 1b. To unfold an unfaulted image, we use non-vertical image flattening (e.g., Luo and Hale, 2011). By unfaulting and then flattening an image, we obtain an image such as the one shown in Figure 1c in which a surface of constant relative geologic time (i.e., a horizontal slice) maps to a geologic horizon in the original image shown in Figure 1a.

2 IMAGE UNFAULTING

To unfault an image, we must first estimate fault locations and fault slip. For the examples shown in this paper, we use the method described by Hale (2012) to automatically compute fault surfaces and fault throws from a 3D seismic image. Although we choose to use Hale’s (2012) method, other methods (e.g., Borgos et al., 2003; Carrillat et al., 2004; Skov et al., 2004; Aurnhammer and Tönnies, 2005; Admasu, 2008; Liang et al., 2010) could also be used to estimate fault locations and fault throw.

For an image $f(x)$, where $x = (x_1, x_2, x_3)$ are coordinates in the present-day space, the estimated fault throw vectors $\mathbf{t}(\mathbf{w})$, where $\mathbf{w} = (w_1, w_2, w_3)$ are coordinates in the unfaulted space, can be used to compute an image

$$\tilde{h}(\mathbf{w}) = f(\mathbf{w} + \mathbf{t}(\mathbf{w}))$$

in which seismic events are aligned across faults where $\mathbf{t}(\mathbf{w})$ is specified. An example of a fault throw vector for a synthetic 2D seismic image is shown in Figure 2. In the figure, $\mathbf{w}_f$ indicates the location of an image sample on the footwall side of the fault, while $\mathbf{w}_h$ indicates the location of the corresponding sample on the hanging wall side of the fault. The fault throw vector $\mathbf{t}(\mathbf{w}_h)$ specifies the location of the image sample that, once shifted to $\mathbf{w}_h$ on the hanging wall side, aligns with the image sample at $\mathbf{w}_f$ on the footwall side. Note from equation 1 that events are shifted only at locations where the fault throw $\mathbf{t}(\mathbf{w})$ is specified. Because we estimate fault throw only at locations where we have identified a fault surface, we must interpolate fault throw vectors at locations between faults to avoid creating new discontinuities in an image when unfaulting.

Our convention is that fault throw vectors $\mathbf{t}(\mathbf{w})$ specify throws on the hanging wall side of a fault. Because we will interpolate these throw vectors (e.g., Figure 4a) between faults, we must also specify fault throw vectors on the footwall side of a fault so that the relative throws on opposing sides of a fault do not change after interpolation. Because fault throw vectors specify throws on only the hanging wall side of a fault, the fault throws on the footwall side must be zero (see Figure 2). For example, Figure 3a shows a subsection of a 3D seismic image from offshore Netherlands with the vertical component of the estimated fault throw vectors overlaid. Notice in Figure 3a that on the hanging wall sides of faults, the fault throws are nonzero, while on the footwall sides, the fault throws are zero. Note that Figure 3a and Figure 4a show the same fault throws on the hanging wall sides of faults, but only Figure 3a shows the zero-valued fault throws on the footwall sides of faults.

To interpolate fault throw vectors at locations between faults, we use blended neighbor interpolation (Hale, 2009a). For Euclidean distances, blended neighbor interpolation is similar to natural neighbor interpolation (Sibson, 1981) and discrete Sibson interpolation (Park et al., 2006). Although we use blended neighbor interpolation, in principle any smooth interpolation would suffice. However, it is important that the interpolation satisfies the interpolation condition, which requires that the interpolant at known locations matches exactly the known values, because we must make sure
Figure 3. A seismic image (a) overlaid with the vertical component of fault throw vectors, and the blended neighbor interpolation (b) of the vertical component of fault throw vectors.
that interpolation does not change the fault throws estimated at fault locations.

To interpolate a vector field, we interpolate separately each vector component. For example, Figure 3b shows the blended neighbor interpolation of the vertical component of the fault throw vectors shown in Figure 3a. Using the interpolated throw vectors $t(w)$ estimated from an input image $f(x)$, the unfaulted image $h(w)$ is computed as

$$h(w) = f(w + t(w)).$$  \hfill (2)

Figure 4b shows the unfaulted image computed according to equation 2 using the interpolated throw vectors shown in Figure 3b and the input image shown in Figure 3a. Similarly, Figure 6b shows the unfaulted image computed from the input image shown in Figure 6a and the blended neighbor interpolation of the fault throw vectors whose vertical component is overlaid on the image in Figure 6a.

3 IMAGE FLATTENING

Automatic image flattening (Stark, 2004; Lomask et al., 2006; Parks, 2010) is a process for computing seismic images in which sedimentary layering is horizontal. Most methods for automatic flattening are limited to only vertical shearing of an image. Flattening by vertical shear, however, can significantly distort image features (Luo and Hale, 2011). Moreover, for seismic images such as the one shown in Figure 4a with non-vertical faults, the true geologic deformation clearly is not vertical. For this reason, we do not limit our flattening method to vertical shear only, but instead allow for non-vertical flattening shift vectors.

To flatten an (unfaulted) image $h(w)$, we must find a mapping $w(u)$, where $u = (u_1, u_2, u_3)$ are coordinates in the flattened space, such that the image

$$g(u) = h(w(u))$$  \hfill (3)

is flat. We write the mapping $w(u)$ in terms of a shift vector field $r(u)$:

$$w(u) = u - r(u),$$  \hfill (4)

which has a corresponding Jacobian matrix $J = \partial w / \partial u$:

$$J = \begin{bmatrix}
1 - \partial r_1 / \partial u_1 & -\partial r_1 / \partial u_2 & -\partial r_1 / \partial u_3 \\
1 - \partial r_2 / \partial u_2 & -\partial r_2 / \partial u_2 & -\partial r_2 / \partial u_3 \\
1 - \partial r_3 / \partial u_3 & -\partial r_3 / \partial u_3 & -\partial r_3 / \partial u_3
\end{bmatrix}. $$  \hfill (5)

Next, given normal vectors $n = (n_1, n_2, n_3)$, which we compute from an image using structure tensors (van Vliet and Verbeek, 1995; Fehmers and Höcker, 2003), we can write the Jacobian matrix for rotation:

$$J_r = \begin{bmatrix}
n_1 + n_2^2/(1 + n_3) & -n_1 n_2/(1 + n_3) & n_1 \\
-n_1 n_2/(1 + n_3) & n_3 + n_1^2/(1 + n_3) & n_2 \\
-n_1 & -n_2 & n_3
\end{bmatrix}. $$  \hfill (6)

Normal vectors transform with the transpose of the Jacobian (Parks, 2010; Luo and Hale, 2011):

$$J^\top n = m,$$  \hfill (7)

where $m$ are normal vectors in the flattened space, and if all normal vectors $m$ point downward in an image, i.e., $m = [0 0 1]^\top$, then the image is flat. It is straightforward to check that $J_r^\top n = [0 0 1]^\top$, so that if $J = J_r$, then applying the shifts $r(u)$ will flatten the image from which the normal vectors $n$ were computed.

To flatten an image, we solve for an approximately isometric mapping $w(u)$ with Jacobian matrix $J$ that satisfies

$$J^\top J_r = I,$$  \hfill (8)

where $I$ is the identity matrix. Isometric mappings are desirable because they preserve metric properties. Thus, if we could isometrically map an image to a flattened image, then all metric properties (e.g., length, angle, area, and volume) of features in the original image would be preserved in the flattened image. Isometric mappings, however, exist only in special cases (Floater and Hormann, 2005), so in general, we solve for a mapping $w(u)$ that is only approximately isometric.

The columns of $J$ contain vectors $w_1(u), w_2(u),$ and $w_3(u).$ That is,

$$J = [w_1(u) \ w_2(u) \ w_3(u)].$$  \hfill (9)

where

$$w_1(u) = \frac{\partial w(u)}{\partial u_1} = \begin{bmatrix} 1 - \partial r_1 / \partial u_1 & -\partial r_2 / \partial u_1 & -\partial r_3 / \partial u_1 \end{bmatrix}^\top,$$

$$w_2(u) = \frac{\partial w(u)}{\partial u_2} = \begin{bmatrix} -\partial r_1 / \partial u_2 & 1 - \partial r_2 / \partial u_2 & -\partial r_3 / \partial u_2 \end{bmatrix}^\top,$$

$$w_3(u) = \frac{\partial w(u)}{\partial u_3} = \begin{bmatrix} -\partial r_1 / \partial u_3 & -\partial r_2 / \partial u_3 & 1 - \partial r_3 / \partial u_3 \end{bmatrix}^\top. $$  \hfill (10)

Vectors $w_1(u)$ and $w_2(u)$ are tangent to a surface (e.g., a horizon) at $u$ and thus are orthogonal to a vector $n(u)$ normal to the surface at $u$ (see Figure 8). The vector $w_3(u)$ is tangent to the line for which the horizontal coordinates $u_1$ and $u_2$ in the flattened space are constant, i.e., the line in coordinates $w$ that maps to a vertical line in coordinates $u$ (Mallet, 2004). For an exactly isometric mapping $w(u),$ tangent vectors $w_1(u), w_2(u),$ and $w_3(u)$ are orthonormal vectors, and the corresponding Jacobian matrix is orthogonal.

Next, if we denote the columns of $J_r$ as $w_1(u), w_2(u),$ and $w_3(u),$ then

$$J_r = [w_1(u) \ w_2(u) \ w_3(u)],$$  \hfill (11)

and equation 8 states

$$\begin{bmatrix} w_1^\top w_1 & w_1^\top w_2 & w_1^\top w_3 \\
w_2^\top w_1 & w_2^\top w_2 & w_2^\top w_3 \\
w_3^\top w_1 & w_3^\top w_2 & w_3^\top w_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \end{bmatrix}. $$.  \hfill (12)
Figure 4. A seismic image (a) overlaid with the vertical component of fault throw vectors, and the unfaulted image (b).
Figure 5. The seismic image shown in Figure 4a is unfaulted and flattened (a) using the composite shift vectors (b).
Figure 6. A seismic image (a) overlaid with the vertical component of fault throw vectors, and the unfaulted image (b).
Figure 7. The seismic image shown in Figure 4a is unfaulted and flattened (a) using the composite shift vectors (b).
The matrix $J^T J$, is a metric tensor characterizing local metric properties such as length, angle, area, and volume (Mallet, 2002, 2004), so by setting this matrix equal to the identity, we constrain the type of deformation parameterized by the mapping $w(u)$. Equation 12 gives nine equations for the partial derivatives of the shift vector field $r(u)$:

$$
n_1 \left(1 - \frac{\partial r_1}{\partial u_1}\right) - n_2 \frac{\partial r_2}{\partial u_1} - n_3 \frac{\partial r_3}{\partial u_1} = 0, \\
-n_1 \frac{\partial r_1}{\partial u_2} + n_2 \left(1 - \frac{\partial r_2}{\partial u_2}\right) - n_3 \frac{\partial r_3}{\partial u_2} = 0, \tag{13}
$$

and

$$
\alpha \left(1 - \frac{\partial r_1}{\partial u_1}\right) - \gamma \frac{\partial r_2}{\partial u_1} + n_1 \frac{\partial r_3}{\partial u_1} = 1, \\
\gamma \left(1 - \frac{\partial r_1}{\partial u_1}\right) - \beta \frac{\partial r_2}{\partial u_1} + n_2 \frac{\partial r_3}{\partial u_1} = 0, \\
-\alpha \frac{\partial r_1}{\partial u_2} + \gamma \left(1 - \frac{\partial r_2}{\partial u_2}\right) + n_1 \frac{\partial r_3}{\partial u_2} = 0, \\
-\gamma \frac{\partial r_1}{\partial u_2} + \beta \left(1 - \frac{\partial r_2}{\partial u_2}\right) + n_2 \frac{\partial r_3}{\partial u_2} = 1, \tag{14}
$$

and

$$
-\alpha \frac{\partial r_1}{\partial u_3} - \gamma \frac{\partial r_2}{\partial u_3} - n_1 \left(1 - \frac{\partial r_3}{\partial u_3}\right) = 0, \\
-\gamma \frac{\partial r_1}{\partial u_3} - \beta \frac{\partial r_2}{\partial u_3} - n_2 \left(1 - \frac{\partial r_3}{\partial u_3}\right) = 0, \\
-n_1 \frac{\partial r_1}{\partial u_3} - n_2 \frac{\partial r_2}{\partial u_3} + n_3 \left(1 - \frac{\partial r_3}{\partial u_3}\right) = 1, \tag{15}
$$

where

$$
\alpha = n_3 + n_2^2/(1 + n_3), \\
\beta = n_3 + n_1^2/(1 + n_3), \\
\gamma = -n_1 n_2/(1 + n_3). \tag{16}
$$

We solve equations 13, 14, and 15 for the components of the shift vector field $r(u)$ by weighted least-squares using conjugate gradient iterations.

Equation 8 describes an isometric mapping of an image to a flattened image, but in general, we cannot expect to find an exactly isometric mapping for all images. In fact, the only image for which we can find an exactly isometric mapping is one in which the normal vectors are constant. For all other images, equations 13, 14, and 15 cannot be satisfied exactly, and we must decide which equations to emphasize.

Equations 13, 14, and 15 correspond to entries in the metric tensor on the left side of equation 12, and they characterize the lengths of and angles between tangent vectors $w_1(u)$, $w_2(u)$, and $w_3(u)$. Specifically, the diagonal entries of the metric tensor characterize lengths of tangent vectors, while the off-diagonal entries characterize angles between tangent vectors. For image flattening, we give most weight to equations 13, which determine the angle between the surface tangent vectors $w_1(u)$ and $w_2(u)$ and the normal vector $n(u)$. If these equations are satisfied, then the image $g(u) = h(w(u)) = h(u - r(u))$ obtained by applying the shifts $r(u)$ will be flat.

We give less weight to equations 14. The four corresponding entries in the metric tensor (equation 12) form what is referred to as the first fundamental form (Floater and Hormann, 2005), which characterizes lengths, areas, and angles measured on a surface. If the first fundamental form equals the identity, then the surface is said to be locally developable, meaning it is isometric to a plane. Because we assume that a geologic horizon was initially deposited as a horizontal layer (i.e., a plane), a developable geologic horizon would indicate that the thickness of sedimentary layers measured perpendicular to bedding has been preserved since its initial deposition.

Finally, we give least weight to equations 15, which determine the length of the tangent vector $w_3(u)$ and the angles it forms with the surface tangent vectors. Recall that $w_3(u)$ is tangent to the line in coordinates $w$ that maps to a vertical line in the flattened coordinates $u$. Thus, if $w_3(u)$ is a unit vector parallel to the normal vector $n$, then the thickness of sedimentary layers measured perpendicular to bedding will be preserved in the flattening process. For most images, we cannot preserve thickness while flattening, so we give the corresponding equations least weight.

Figures 5a and 7a show flattened unfauluted images computed from the unfaulted images shown in Figures 4a and 6a, respectively. For both examples, normal vectors were computed using structure tensors with Gaussian smoothing filters (Hale, 2009b) with a vertical half-width of 32 ms and horizontal half-widths of 50 m. In addition, when solving equations 13, 14, and 15, we used...
smoothing preconditioners as described by Parks (2010) to speed convergence. The smoothing filter in the vertical direction had a half-width of 24 ms, while the filters in the horizontal directions had half-widths of 150 m.

4 HORIZON EXTRACTION

Once we obtain a flattened unfaulted image, the corresponding flattening shift vectors, and the interpolated throw vectors, we can extract geologic horizons such as those shown in Figure 9.

We extract a horizon by first selecting a horizontal slice of constant \( u_3 \) in a flattened unfaulted image \( g(u) \). Next we form a composite mapping \( x(u) \) by combining the mapping \( x(w) = w + t(w) \) used to unfault an image with the mapping \( w(u) = u - r(u) \) used to flatten an image to obtain

\[
x(u) = u - s(u),
\]

where \( s(u) \) is the composite shift vector field:

\[
s(u) = r(u) - t(u - r(u)),
\]

which allows for a direct mapping from an image \( f(x) \) to a flattened unfaulted image \( g(u) \) with

\[
g(u) = f(u - s(u)).
\]

Figures 9a and 9b show geologic horizon surfaces extracted from the composite shift vector fields shown in Figures 5b and 7b, respectively. In the horizon in Figure 9a, notice the en échelon faults that can be clearly seen in the seismic image in Figure 4a. In the horizon in Figure 9b, notice the roughly circular fault polygons, which correspond to the conical fault surfaces described by Hale (2012). Note that although we show only a single horizon for each image, it is trivial to extract another horizon by simply choosing a different horizontal slice of constant \( u_3 \) from a flattened unfaulted image.

5 CONCLUSION

We have presented a method to automatically unfault and flatten seismic images. The method requires an estimate of fault locations and fault throw vectors, which, for the examples shown in this paper, we obtain using the method described by Hale (2012). Flattened unfaulted images are images in which horizontal slices correspond to a constant geologic time, and the geologic age of horizontal slices increases with increasing time.

The flattening method we describe solves, in a least-squares sense, for an approximately isometric mapping of an image to a flattened image. For such a mapping, it is assumed that geologic folding follows the flexural slip style of deformation (Suppe, 1985) with constant sedimentation velocity (Mallet, 2002). This assumption might not always be correct, but we can evaluate it using the composite shift vector field used to compute a flattened unfaulted image. For example, we can compute the tangent vectors given in equation 9 to determine if they are orthonormal. Also, we can compute properties of geologic horizons such as Gaussian curvature to
determine which horizons are locally developable. Note that these computations require partial derivatives of the mapping \( x(u) = u - s(u) \). For images in which geologic horizons are horizontal (i.e., flattened unfaulted images), these derivatives can be computed with a simple finite-difference scheme, but for images in which horizons are deformed, derivatives are less straightforward to compute because they require knowledge of the surface parameterization.

Some limitations remain in our method. One is that we are currently using only estimated fault throw vectors to unfault an image. In reality, fault slip consists of fault heave as well as fault throw. Fault heave, however, is more difficult to estimate not only because fault heave tends to be parallel to sedimentary layering, but also because seismic images have lower lateral than vertical resolution. Another limitation arises from the way in which we compute normal vectors. Because normal vectors are computed in local windows (and, moreover, are constrained to point in the positive time or depth direction), they cannot distinguish overturned or recumbent horizons. This limitation could be overcome by using a different method for computing normal vectors or by filtering normal vectors computed using structure tensors to distinguish overturned or recumbent horizons.

Although the method described in this paper consists of two steps—image unfaulting followed by image flattening—it is possible to combine these two steps into one, by solving for shifts that flatten an image while constraining the shift vectors to be equal to fault throw vectors estimated at fault locations. Recall that when flattening, we try to preserve surface metric tensors (equation 14) and sedimentary layer thickness (equation 15), while emphasizing image flatness (equation 13). The flattening process, however, can only preserve surface metric tensors and layer thickness with respect to the input image, which, for the examples shown in this paper, was an unfaulted image. Preserving surface metric tensors and layer thickness with respect to the original faulted and folded image would be more appropriate, but to do so would require the one-step image unfaulting and flattening process described above.

ACKNOWLEDGMENTS

Thanks to dGB Earth Sciences for providing the 3D seismic images shown in this report. This research was supported by the sponsors of the Center for Wave Phenomena at the Colorado School of Mines.

REFERENCES


———, 2004, Relative geologic time (age) volumes — Relating every seismic sample to a geologically reasonable horizon: The Leading Edge, 23, 928–932.
Dynamic warping of seismic images

Dave Hale

Center for Wave Phenomena, Colorado School of Mines, Golden CO 80401, USA

ABSTRACT

The problem of estimating relative time (or depth) shifts between two seismic images is ubiquitous in seismic data processing. This problem is especially difficult where shifts are large and vary rapidly with time and space, and where images are contaminated with noise. I propose a solution to this problem that is a simple extension of the classic dynamic time warping algorithm for speech recognition. This new dynamic image warping method for estimating shifts is more accurate than methods based on crosscorrelation of windowed images where shifts vary significantly within the windows.

Key words: seismic image dynamic warping correlation

1 INTRODUCTION

In seismic data processing we must often estimate relative shifts in time (or depth) between seismograms. Often those shifts vary with both time and space coordinates. Examples cited by Liner and Clapp (2004) include alignment of synthetic and recorded seismograms, registration of P- and S-wave images, residual normal moveout correction, and alignment of images computed for different source-receiver offsets or propagation angles. They proposed a dynamic programming solution to this problem in the case where pairwise alignment between seismic traces is sufficient.

A different dynamic programming solution was developed by Sakoe and Chiba (1978) in the context of speech recognition, and is today widely known as dynamic time warping (DTW; e.g., Müller, 2007, Chapter 4). Significantly, DTW imposes constraints on the rate at which shifts may vary in time, and these constraints often enable DTW to accurately estimate shifts from sequences that are contaminated with noise, or that in some other ways are not simply warped versions of each other.

The use of dynamic time warping to estimate shifts in geophysical time series and other sequences is not new. Several applications of dynamic time warping to problems in geophysics were proposed by Anderson and Gaby (1983), who called this algorithm “dynamic waveform matching”.

Unfortunately, the most straightforward extension of DTW to the problem of estimating shifts in multidimensional images has been shown to be NP-complete
(Keysers and Unger, 2003), that is, computationally intractable. Therefore, numerous authors — Pishchulin (2010) provides a recent summary — have proposed practical solutions to problems that approximate this NP-complete problem.

In this paper I propose an extension of an approximate solution developed by Mottl et al. (2002). Their solution and my extension for dynamic image warping (DIW) are especially simple, requiring very little software beyond what would already be available to implement dynamic time warping. I provide the essential computer software components for both DTW and DIW in Appendix A.

I first review the dynamic time warping algorithm, giving special attention to the so-called accumulation and backtracking parts of this algorithm. I then show how the accumulation part of DTW can be used to implement a non-linear smoothing of alignment errors computed for two multi-dimensional images, and how this leads to a new method for DIW.

In tests with pairs of images related by shifts that are known, large, and rapidly varying, I demonstrate the accuracy with which DIW can estimate the known shifts. Figure 1 is an example of one such test for two images contaminated with bandlimited random noise.

In further tests I show that DIW can be more accurate than methods based on local crosscorrelations, especially when shifts vary rapidly in time or space. Crosscorrelation methods, such as those proposed by Hall (2006) and Hale (2009) to estimate shifts in time-lapse seismic images, are accurate only where shifts are more slowly varying.

2 DYNAMIC TIME WARPING

Consider the two synthetic seismograms \( f[i] \) and \( g[i] \) with length \( N = 500 \) samples displayed in Figure 2. I computed the sequence \( f[i] \) by convolving a Ricker wavelet with a random reflectivity sequence. I then applied time-varying shifts \( s[i] \) to that reflectivity sequence and convolved again with the same wavelet to obtain the sequence \( g[i] \). The two sequences are therefore approximately (not exactly) related by \( f[i] \approx g[i + s[i]] \).

In practice these two sequences might be a recorded seismogram \( f[i] \) and a synthetic seismogram \( g[i] \) derived from well logs. Or they might be two sequences of sample values extracted from a seismic image on opposite sides of a fault. In any case, the practical problem considered here is estimation of the shifts \( s[i] \) given only the two sequences \( f[i] \) and \( g[i] \).

In the example of Figure 2, the shifts \( s[i] \) are a simple sinusoidal function that is apparent in Figure 3a, which is an image of alignment errors defined by

\[
e[i, l] = (f[i] - g[i + l])^2.
\]

Note that these alignment errors are nearly zero where the integer lag \( l \) approximately equals the shift \( s[i] \). Also note the constant extrapolation of errors in the corners of \( e[i, l] \), where \( i + l < 0 \) or \( i + l > N \).

The definition of alignment errors in equation 1 can be modified by changing the power 2 or by using some other non-negative function of the differences \( f[i] - g[i + l] \), without changing the dynamic time warping algorithm. For example, we might use the absolute values of those differences. In all of the examples shown in this paper I have simply squared the differences, as in equation 1.
2.1 Constrained optimization

The simplest dynamic time warping (DTW) computes a sequence \( u[0 : N – 1] \equiv (u[0], u[1], \ldots, u[N – 1]) \) of integer shifts by solving the following optimization problem:

\[
u[0 : N – 1] = \arg \min_{l[0 : N – 1]} D(l[0 : N – 1]), \tag{2}\]

where

\[
D(l[0 : N – 1]) \equiv \sum_{i=0}^{N-1} e[i, l[i]], \tag{3}
\]

subject to the constraint

\[
|u[i] – u[i-1]| \leq 1. \tag{4}
\]

As illustrated in Figure 3b, DTW yields a minimizing sequence of integer shifts \( u[0 : N – 1] \) that well approximates the (here, known) sequence of shifts \( s[0 : N – 1] \).

The function \( D \) defined by equation 3 is often referred to as distance, which makes sense if we think of the image \( e[i, l] \) in Figure 3a as representing some topography. Large values in \( e[i, l] \) then correspond to tall hills (misalignments) that we wish to avoid as we choose a path from left to right, that is, from \( i = 0 \) to \( N – 1 \). In this sense, DTW chooses a path \( u[0 : N – 1] \) to minimize the total distance traveled, subject to the constraint (equation 4) that the shifts cannot change too rapidly from one sample to the next.

The constraint equation 4 is analogous to the simplest slope constraint of Sakoe and Chiba (1978). This constraint is slightly different here only because I define alignment errors by \( e[i, l] \equiv (f[i] – g[i+l])^2 \) (equation 1), instead of by \( e[i, j] \equiv (f[i] – g[j])^2 \). Here this constraint ensures that the argument \( i + u[i] \) in \( (f[i] – g[i + u[i]])^2 \) neither decreases nor increases too rapidly with increasing sample index \( i \).

This constraint is important. Where \( u[i] – u[i-1] = 1 \), we stretch by 100%, such that two adjacent samples in the sequence \( f[i] \) correspond to two non-adjacent samples in the sequence \( g[i] \). Where \( u[i] – u[i-1] = -1 \), we squeeze by 100%, such that two adjacent samples in the sequence \( f[i] \) correspond to only one sample in the sequence \( g[i] \). In many practical applications 100% is an unreasonably large amount of strain, and we will see below how to reduce this upper bound.

It is significant that the sequence \( u[0 : N – 1] \) computed by DTW minimizes exactly the distance \( D \) defined by equation 3, while satisfying the constraint equation 4. Differences in Figure 3b between the integer shifts \( u[0 : N – 1] \) and the known shifts \( s[0 : N – 1] \) are due entirely to the restriction of \( u[0 : N – 1] \) to be integers and the approximation \( f[i] \approx g[i + s[i]] \), not to any approximation in the optimization algorithm.

2.2 Dynamic programming

As its name implies, dynamic time warping is a dynamic-programming algorithm (e.g., Cormen et al., 2001). The essential trait of this algorithm is decomposition of a problem into a sequence of nested and smaller subproblems.

Let \( u[0 : N – 1] \) denote the sequence of shifts \( l \) that minimizes the distance \( D \) defined by equation 3. To identify the sequence of smaller subproblems nested within this larger minimization problem, we consider a subpath \( u[0 : m] \) of the minimizing path \( u[0 : N – 1] \) and observe that

\[
u[0 : m] = \arg \min_{l[0 : m]} \sum_{i=0}^{j} e[i, l[i]]. \tag{5}\]

For if \( u[0 : m] \) were not a minimizing subpath, then we could replace that part of \( u[0 : N – 1] \) and thereby reduce the total distance \( D \), which implies that \( u[0 : N – 1] \) does not minimize \( D \), a contradiction.

This observation is important because it implies that we need not test all possible (roughly, \( 3^N \)) paths \( l[0 : N – 1] \) that satisfy the constraint equation 4 in our search for the minimizing path \( u[0 : N – 1] \). Instead, we can find this minimizing path in two steps: accumulation and backtracking.

2.3 Accumulation

In the first accumulation step of DTW, we recursively compute from the array of alignment errors \( e[i, l] \) an array of distances \( d[i, l] \) as follows:

\[
d[0, l] = e[0, l],
\]

\[
d[i, l] = e[i, l] + \min \left\{ d[i-1, l-1], d[i-1, l], d[i-1, l+1] \right\},
\]

for \( i = 1, 2, \ldots, N – 1 \). \tag{6}

For each index \( i \), we cannot yet know in this first step whether or not the lag \( l \) lies on the minimizing path \( u[0 : N – 1] \), so that \( l = u[i] \). Therefore, we must compute and store distances \( d[i, l] \) for all lags, assuming for the moment that lag \( l \) may lie on the minimizing path. Figure 3b shows distances \( d[i, l] \) computed in this way for the alignment errors shown in Figure 3a.

The constraint equation 4 implies that, when computing \( d[i, l] \) as in equation 6, we must consider only three previously computed distances \( d[i-1, l-1], d[i-1, l], \) and \( d[i-1, l+1] \). In other words, if lag \( l \) lies on the minimizing path at sample index \( i \), then either lag \( l-1 \), \( l \), or \( l+1 \) must lie on the minimizing path at sample index \( i-1 \).

The computational complexity of this first step is \( O(N \times L) \), where \( L \) is the number of lags \( l \) for which alignment errors \( e[i, l] \) and distances \( d[i, l] \) are computed.

I call this first step accumulation because the distances \( d[i, l] \) are sums of alignment errors \( e[i, l] \). At the end of this first step, we can simply loop over all lags \( l \).
The robustness of DTW in the presence of random noise

3.1 Limiting strain

Figure 4a.

is not simply a warped version of the sequence

dition of random noise. The sequence

approximation

ferences in Figure 5b between esti-

ergies exactly the constrained optimization problem of

us

plot of alignment errors

obscures somewhat the sinuosidal warping path in the

The rms signal:noise ratio is 2:1. While this level of noise

sequences of bandlimited random noise to each of them.

Figure 4 displays the same two synthetic seismograms

and ending with the first shift $u[0]$:

$u[N - 1] = \arg \min_l d[N - 1, l],$

$u[i - 1] = \arg \min_{l \in \{u[i - 1], u[i], u[i] + 1\}} d(i - 1, l),$

for $i = N - 1, N - 2, \ldots, 1.$

This backtracking step begins with a simple loop over

lags $l$ to find the last shift $u[N - 1]$ in the sequence of

shifts $u[0 : N - 1].$ Because this last shift must be on

the minimizing path, it must equal the lag at which we

found the minimum distance $D.$

We then recursively find previous shifts $u[i - 1]$ in

this sequence, comparing the three distances $d[i - 1, l -

1], d[i - 1, l]$ and $d[i - 1, l + 1]$ to determine which of

these was used in equation 6 to compute the minimum

distance $d[i, l].$

The computational complexity of backtracking is

only $O(N),$ because for each sample index $i$ we compute

a shift $u[i]$ by comparing only three distances. Therefore,

the complexity of DTW is $O(N \times L),$ that of the

accumulation step, which is proportional to the number

of samples in the image of alignment errors $e[i, l].$

3 REFINEMENTS

Figure 4 displays the same two synthetic seismograms

$f[i]$ and $g[i]$ shown in Figure 2, after adding different

sequences of bandlimited random noise to each of them.

The rms signal:noise ratio is 2:1. While this level of noise

obscures somewhat the sinuosidal warping path in the

plot of alignment errors $e[i, l]$ in Figure 5a, the shifts

$u[i]$ estimated using DTW are a rough approximation

to the known shifts $s[i].$

Again it is important to remember that DTW solves

exactly the constrained optimization problem of

equations 2–4. Differences in Figure 5b between esti-
mated and known shifts are primarily due to errors in

the approximation $f[i] \approx g[i + s[i]]$ caused by the

addition of random noise. The sequence $g[i]$ in Figure 4b

is not simply a warped version of the sequence $f[i]$ in

Figure 4a.

3.1 Limiting strain

The robustness of DTW in the presence of random noise

is due largely to the constraint equation 4. The number

of shift sequences $u[0 : N - 1]$ that satisfy this con-

straint ($\approx 3^N$) is far less than the number that would

be possible without it ($\approx L^N$).

Of course, the constraint that strain (stretch or

squeeze) be less than 100% is useful only when satisfied

by the actual shifts $s[i]$ that we wish to estimate. How-

ever, in many practical applications this constraint is

more than reasonable. Indeed, a strain as high as 100%

may be unreasonably high, and we may be able to im-

prove the accuracy of shifts estimated in DTW by re-

ducing this upper bound on strain to a more reasonable

value.

The simplest way to more tightly bound strain in

DTW is to sample lags $l$ more finely at some fraction

of the time sampling interval. For example, if that frac-
tion were \( \frac{1}{2} \), then we would compute alignment errors \( e[i, l] \) for lags \( l = \ldots, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, \ldots \). The maximum strain permitted would then be 50%, as the constraint equation 4 would become

\[
|u[i] - u[i - 1]| \leq \frac{1}{2}
\]  
(9)

While straightforward, this method for reducing the upper bound on strain requires a significant increase in computational cost. For a limit of 50%, both computation time and memory will double if we compute errors \( e[i, l] \) and distances \( d[i, l] \) for twice as many lags. The increase in memory will be especially significant as we extend the dynamic time warping algorithm to the problem of multi-dimensional image warping.

A more efficient way to limit strain is to implement constraints much like the slope constraints proposed by Sakoe and Chiba (1978). As an example, for a limit of 50% strain, I change the accumulation step (equation 6) to compute distances as

\[
d[0, l] = e[0, l],
\]

\[
d[1, l] = e[1, l] + \min\left\{d[0, l - 1], d[0, l], d[0, l + 1]\right\},
\]

\[
d[i, l] = e[i, l] + \min\left\{d[i - 2, l - 1] + e[i - 1, l - 1], d[i - 1, l - 1], d[i - 2, l + 1] + e[i - 1, l + 1]\right\}
\]  
for \( i = 2, 3, \ldots, N - 1 \).  
(10)

A corresponding change is required in the backtracking step, in which we must now compute and compare the three expressions inside the min function of equation 10, to determine which of these was used to compute the distance \( d[i, l] \).

The effect of these modifications is to impose the following constraint on changes in shifts:

\[
|u[i] - u[i - 1]| + |u[i - 1] - u[i - 2]| \leq 1.
\]  
(11)

Like equation 9, this equation is similar to (but not quite equivalent to) a finite-difference approximation to \( |du/dt| \leq \frac{1}{2} \).

In words, dynamic time warping based on equation 10 is constrained to shift sequences in blocks of two or more samples. If any sample is shifted by the warping, then at least one of the adjacent samples must be shifted by the same amount.

Modifications similar to equation 10 can be easily and efficiently implemented for any strain limit of the form \( 1/b \), where \( b \) is an integer. (See Appendix A.) Figure 6 shows how strain limits implemented in this way can improve the accuracy of shifts estimated by DTW.

Note, however, that the strain limit of \( \frac{1}{2} \) used to estimate shifts \( u[i] \) shown in Figure 6c is almost equal to the maximum strain in the known shifts \( s[i] \). Any further reduction in the strain limit would yield poor shift estimates, because strain for the correct shifts would exceed that limit. In practice, lacking any a priori limit on strain, we must take care to not reduce this limit so much that we prohibit the correct shifts.

### 3.2 Smoothing alignment errors

To further improve the accuracy of estimated shifts \( u[i] \), we might attempt to attenuate noise in the two sequences \( f[i] \) and \( g[i] \), or we might instead try to attenuate noise in the alignment errors \( e[i, l] \). Considering the second option, suppose that we apply some sort of smoothing filter to the alignment errors \( e[i, l] \). Can we improve the accuracy of the estimated shifts \( u[i] \) by applying such a filter before DTW?

This question is suggested by the accumulation step in DTW defined by equation 6. Each distance \( d[i, l] \) computed in this step is a sum of alignment errors, which implies that the distances \( d[i, l] \) vary less rapidly with index \( i \) than do the alignment errors \( e[i, l] \). In other words, the accumulation step is a smoothing filter.

This recursive smoothing filter is one-sided, because each \( d[i, l] \) in equation 6 depends on only previous and present alignment errors, those with sample indices less than or equal to \( i \). This filter is also non-linear, because of the min function in equation 6. In effect, this one-sided non-linear smoothing filter already attenuates...
noise in alignment errors caused by noise in the two sequences to be aligned by warping.

One way we might improve this smoothing filter would be to make it two-sided and symmetric. We can implement a two-sided symmetric smoothing filter by applying a one-sided filter in forward and reverse directions. Smoothing in the forward direction is the same as computing distances \( d[i, l] \) via equation 6:
\[
\tilde{e}_f[i, l] = e[i, l] + \begin{cases} \tilde{e}_f[i - 1, l - 1] \\ \tilde{e}_f[i - 1, l] \\ \tilde{e}_f[i - 1, l + 1] \end{cases}
\]
for \( i = 1, 2, \ldots, N - 1 \).

(12)

Smoothing in the reverse direction is similar:
\[
\tilde{e}_r[N - 1, l] = e[N - 1, l],
\tilde{e}_r[i, l] = e[i, l] + \begin{cases} \tilde{e}_r[i + 1, l - 1] \\ \tilde{e}_r[i + 1, l] \\ \tilde{e}_r[i + 1, l + 1] \end{cases}
\]
for \( i = N - 2, N - 3, \ldots, 0 \).

(13)

Two-sided smoothing is then defined by
\[
\tilde{e}[i, l] = \tilde{e}_f[i, l] + \tilde{e}_r[i, l] - e[i, l].
\]

(14)

Subtraction of \( e[i, l] \) in equation 14 ensures that this value is not counted twice, as for all \( i \) it appears in both \( \tilde{e}_f[i, l] \) and \( \tilde{e}_r[i, l] \). In this way, each smoothed error \( \tilde{e}[i, l] \) is a sum of past, present and future alignment errors.

Like the accumulator in DTW, this two-sided smoothing filter is non-linear because it uses the min function in both equations 12 and 13 to determine which errors to sum. Figure 7 displays smoothed alignment errors for the two noisy sequences shown in Figure 4.

Observe that the known sinusoidal warping path is somewhat more apparent in these smoothed errors than in the unsmoothed alignment errors displayed in Figure 5a. We might therefore expect the shifts \( u[i] \) estimated by DTW from the smoothed errors \( \tilde{e}[i, l] \) would be more accurate than those estimated from the unsmoothed errors \( e[i, l] \).

However, this sort of smoothing does not improve DTW. Although not shown here, the shifts \( u[i] \) computed by DTW for the smoothed alignment errors shown in Figure 7c are identical to those computed for the unsmoothed alignment errors in Figure 5a. The benefit of this two-sided smoothing lies in the extension of dynamic warping for one-dimensional sequences to that for multi-dimensional images.

4 DYNAMIC IMAGE WARPING

The simplest way to extend dynamic time warping for image processing is to think of an image as a collection of vertical columns and to estimate vertical shifts by applying DTW to each of those columns independently. We could likewise apply DTW to image rows to obtain estimates of horizontal shifts.

Figure 8 illustrates the application of this simple method for dynamic image warping for two seismic shot records, where the first record shown in Figure 8a has been warped to obtain the second record shown in Figure 8b. DTW applied to each corresponding pair of columns from these images yields the estimated shifts shown in Figure 8a. Except for small source-receiver offsets where seismograms are missing, the estimated shifts approximate well the known shifts used to warp the images.

These shifts are large, about eight times larger than the dominant period of most reflection events, which is about 40 ms. And many of the events in the shot records (such as those for small offsets and late times) are ringy, almost periodic, which can make estimation of the shifts more difficult. Nevertheless, DTW applied independently to each pair of seismograms in these shot records is able to recover the correct shifts.

The success of this simple method for dynamic image warping (DIW) depends on the fact that each pair of seismograms in these shot records satisfies exactly the DTW assumption that one sequence is a warped version of the other.

When this assumption is not satisfied, this simple
method for DIW can fail miserably. For example, if we add different bandlimited random noise images to each of the shot records before DIW, we obtain the results shown in Figure 9. In this example, the rms signal-to-noise ratio is 1:1. For this noise level, it is difficult to estimate well the correct shifts from each pair of noisy seismograms in the two shot records. Therefore, the estimated shifts vary significantly for different offsets, and imply an unlimited amount of strain in the horizontal direction.

To improve these estimated shifts, we would like to limit strain in both horizontal and vertical directions. In other words, we would like to minimize alignment errors as in equations 2 and 3 while satisfying constraints like those in equations 4 or 11 in both horizontal and vertical directions.

Unfortunately, this constrained optimization problem has been shown to be NP-complete (Keysers and Unger, 2003), which means that the existence of a computationally feasible solution is highly unlikely. We must therefore make approximations to this NP-complete problem, and methods for DIW differ in their approximations.

### 4.1 Tree-sequential dynamic programming

One such approximation is that proposed by Mottl et al. (2002), and this approximation and its solution are today often referred to as tree-sequential dynamic programming (TSDP; e.g., Pishchulin, 2010).

Given software for DTW like that in Appendix A, implementation of the TSDP algorithm for DIW is almost trivial in the case considered here, where we seek to estimate only vertical shifts. The TSDP algorithm begins by computing alignment errors as for DTW. It then smooths those alignment errors in the vertical direction, by applying the non-linear two-sided smoothing equations 12–14 independently for each image column. TSDP ends by applying the DTW algorithm to the smoothed errors, but now in the horizontal direction, accumulating and backtracking for each image row, again independently. In this way, TSDP processes a multi-
Figure 10. Time shifts estimated by dynamic time warping after vertical (a), vertical-horizontal (b), vertical-horizontal-vertical (c), and vertical-horizontal-vertical-horizontal (d) smoothings of alignment errors. Insignificant differences between shifts (c) and (d) indicate that this process of smoothing in alternating directions before DTW has converged.

dimensional image with a cascade of one-dimensional smoothing, accumulation and backtracking.

Time shifts estimated by TSDP are shown in Figure 10a. For this example I used strain limits of 25% in the vertical direction and 100% in the horizontal direction, and these values are close to the maximum strains in the known shifts displayed in Figure 9d. Compared to the estimated shifts shown in Figure 9c, the shifts from TSDP shown in Figure 10a better approximate the known shifts.

The most obvious improvement is in reduced horizontal strain, greater continuity of shifts in the horizontal direction. This improvement is not surprising, because TSDP as described above ends by applying DTW independently for each image row, and we know that DTW satisfies strain limits precisely.

However, because TSDP ends by applying DTW independently for each row, we have no guarantee that vertical strain limits (here 25%) will be satisfied. Vertical discontinuities in shifts are in fact apparent in Figure 10a. Although vertical smoothing of alignment errors in TSDP reduces the likelihood that such discontinuities will occur, it does not entirely eliminate them.

Instead of smoothing vertically and then applying DTW horizontally, we might instead smooth horizontally and apply DTW vertically. Shifts estimated by this alternative implementation of TSDP are not shown here, but are significantly less accurate than those shown in Figure 10a.

The reason to first smooth vertically is that, for each offset, a pair of image columns (seismograms) typically contains multiple events that will indicate a path of minimum alignment error like that apparent in Figure 3a, but the same is not true for each pair of image rows. By first smoothing alignment errors vertically, we extend these paths of minimum error to times for which little information about vertical alignment may be available, so that DTW applied horizontally can then accurately estimate the shifts. Nevertheless, vertical discontinuities apparent in the estimated shifts shown in Figure 10a suggest that further improvement is possible.

4.2 Improving TSDP

The key to improving TSDP lies in recognizing that it first smooths alignment errors in one direction before it applies DTW in another direction. Although I have not seen TSDP described in this way, the description is accurate. So why not first smooth in both vertical and horizontal directions?

Figure 10b shows the result of smoothing both vertically and horizontally before applying DTW vertically to each column of smoothed alignment errors. As expected, vertical discontinuities in shifts are now eliminated, but a few horizontal discontinuities are apparent. However, if we apply more vertical and horizontal smoothings, this process quickly converges to the smooth shifts shown in Figure 10c and 10d.

Although I have no guarantee that this smoothing process will converge, I have not found a practical example in which more than four (vertical-horizontal-vertical-horizontal) smoothings yielded any significant changes in shifts. The convergence shown in Figure 10 is typical.

Even assuming that the smoothing process does converge, we cannot guarantee that estimated shifts will minimize alignment errors while satisfying both vertical and horizontal strain limits, as we recall that this constrained optimization problem is NP-complete (Keysers and Unger, 2003).

The new dynamic image warping method proposed here is truly an extension of the TSDP method proposed by Mottl et al. (2002). Indeed, one way to view this new method is that it is TSDP with a larger tree, in which each vertical or horizontal smoothing before dynamic time warping represents a new set of branches.
4.3 Dynamic warping and crosscorrelation

In tests of dynamic image warping discussed above, shifts are large (much larger than the dominant period of reflections) and vary rapidly with both time and offset. Recalling that strain is the rate at which shift changes, the maximum strain in time is about 25%, and the maximum strain in offset is almost 100%. That is, time shifts change by as much as a quarter of one time sample from one sampled time to the next, and by almost one time sample from one sampled offset to the next.

Where shifts are not so rapidly varying, methods based on local crosscorrelation of images may be used instead to obtain accurate shift estimates.

Figure 11 displays shifts estimated from noisy shot records like those in Figure 9a and 9b, using both dynamic image warping and local crosscorrelations. The local crosscorrelation method used here is that described by Hale (2009), which finds shifts that maximize correlation coefficients computed for seamlessly overlapping windows of images. In these tests those windows are Gaussian with half-widths equal to 320 ms in time and 240 m in offset.

Figure 11 illustrates how the success of this crosscorrelation method depends on whether or not shifts vary rapidly within the windows used to compute the correlation coefficients. The sinusoidal pattern of variation used for these tests is the same as that shown in Figure 8d, but the rates at which shifts change with time and offset (the strains) are smaller because the magnitudes of the shifts are smaller.

Where shifts vary slowly, as in Figures 11a and 11b (where the maximum time shift is less than four samples), both dynamic image warping and local crosscorrelation yield estimated shifts that approximate well the known shifts. Shifts estimated using the crosscorrelation method show significant errors only for small times and large offsets where no reflections exist.

Where shifts vary rapidly, as in Figures 11c and 11d, (ten times more rapidly than for Figures 11a and 11b), shifts estimated using the local crosscorrelation method are unstable and inaccurate, while those obtained by dynamic image warping again approximate well the known shifts.

One way to stabilize shifts estimated in crosscorrelation methods is maximize a weighted sum of both image correlation and shift smoothness. Hall (2006), for example, used such a regularization to improve the stability and accuracy of small shifts estimated from time-lapse seismic images.

Regularization is however not the same as constraints in DTW, and it does not solve the fundamental problem in using crosscorrelation windows to estimate shifts that vary rapidly within those windows. For such shifts, correlation coefficients will be low for all lags, because the two windowed images cannot be well aligned for any one of them. If we try to solve this problem by using smaller windows, then noise will degrade our shift estimates, and in any case we must use correlation windows that are at least as large as the shifts we seek to estimate. In summary, local correlation methods require that we choose a window size; for shifts that vary rapidly, as for Figure 11d, a suitable choice may not exist.

In contrast, dynamic time and image warping require no windows. We need not specify a window size when using the dynamic image warping algorithm proposed in this paper.

Another difference between crosscorrelation and dynamic warping methods is that correlation peaks are easily found with sub-sample precision, but dynamic warping yields only integer shifts. For this reason I smooth the integer shifts estimated with dynamic warping using Gaussian filters with half-widths equal to image sampling intervals divided by maximum strains used to constrain the shifts.
CONCLUSION

An appealing feature of dynamic time warping is that it solves exactly the constrained optimization problem of equations 2–4. Although a practical and exact solution is unlikely to exist for the corresponding optimization problem for images, examples shown above indicate that the approximate dynamic warping solution proposed in this paper can produce reasonable estimates of shifts from noisy images, even where those shifts are large and rapidly varying.

The examples shown above are for 2D images, but dynamic image warping is easy to apply to 3D images as well. For 3D images we simply smooth alignment errors along all three image dimensions before eventually using dynamic time warping to estimate the shifts. For each dimension, this smoothing and dynamic time warping is especially efficient on computers with multiple processors, as each row or column of alignment errors can be processed in parallel.

In this paper I have assumed that images can be aligned with only vertical warping. However, the dynamic image warping algorithm proposed here can also be used to estimate shift vectors with vertical and horizontal components. Estimating shift vectors would require modification of the software in Appendix A to process arrays of alignment errors computed for multiple components of lag, but these modifications are straightforward.

ACKNOWLEDGMENTS

I am grateful to Luming Liang for emphasizing to me the shortcomings of crosscorrelation methods in estimating shifts that vary rapidly. The seismic shot record used to demonstrate dynamic image warping is one of forty shot records gathered and made freely available by Oz Yilmaz.

REFERENCES


APPENDIX A: SOFTWARE

Listed below are Java methods for computing an array of accumulated alignment errors $d[i, l]$ and for backtracking within such an array to find shifts $u[0 : N - 1]$ that minimize accumulated errors. The accumulate method can be used in both positive and negative directions to compute smoothed alignment errors $\tilde{e}[i, l]$ in dynamic image warping.

Both methods can be translated easily to equivalent functions written in the C or C++ programming languages.
/**
 * Non-linear accumulation of alignment errors.
 * @param dir accumulation direction, + or -.
 * @param b strain is bounded by 1/b.
 * @param e input array of alignment errors.
 * @param d output array of accumulated errors.
 */

void accumulate(int dir, int b,
    float[][] e, float[][] d)
{
    int nl = e[0].length;
    int ni = e.length;
    int nlm1 = nl-1;
    int nim1 = ni-1;
    int ib = (dir>0)?0:nim1;
    int ie = (dir>0)?ni:-1;
    int is = (dir>0)?1:-1;
    for (int i=ib; i!=ie; i+=is) {
        int ji = max(0,min(nim1,i-is));
        int jb = max(0,min(nim1,i-is*b));
        float dm = d[jb][lm1];
        float di = d[ji][l ];
        float dp = d[jb][lp1];
        for (int l=0; l<nl; ++l) {
            int lm1 = l-1; if (lm1<=-1) lm1 = 0;
            int lp1 = l+1; if (lp1==nl) lp1 = nlm1;
            float dl = d[jl][l ];
            float dm += e[jb][lm1];
            float dp += e[jb][lp1];
            float di += e[jl][l ];
            d[i][l] = min3(dm,di,dp)*e[i][l];
        }
    }
}

/**
 * Finds shifts by backtracking within an
 * array of accumulated alignment errors.
 * Backtracking must be performed in the
 * direction opposite that in which the
 * alignment errors were accumulated.
 * @param dir backtrack direction, + or -.
 * @param b strain is bounded by 1/b.
 * @param lmin lag for lag index 0.
 * @param e input array of alignment errors.
 * @param d output array of accumulated errors.
 * @param u output array of shifts.
 */

void backtrack(int dir, int b, int lmin,
    float[][] d, float[][] e, float[] u)
{
    int nl = d[0].length;
    int ni = d.length;
    int nlm1 = nl-1;
    int nim1 = ni-1;
    int ib = (dir>0)?0:nim1;
    int ie = (dir>0)?nim1:0;
    int is = (dir>0)?1:-1;
    float dl = d[il][l];
    for (int jl=1; jl<nl; ++jl) {
        if (d[il][jl]<dl) {
            dl = d[il][jl];
            il = jl;
        }
    }
    u[il] = il+lmin;
    while (il!=ie) {
        int ji = max(0,min(nim1,il+is));
        int jb = max(0,min(nim1,il+is*b));
        int ilm1 = il-1; if (ilm1<=-1) ilm1 = 0;
        int ilp1 = il+1; if (ilp1==nl) ilp1 = nlm1;
        float dm = d[jb][ilm1];
        float di = d[ji][il ];
        float dp = d[jb][ilp1];
        for (int kb=ji; kb!=jb; kb+=is) {
            dm += e[kb][ilm1];
            dp += e[kb][ilp1];
        }
        dl = min3(dm,di,dp);
        if (dl!=di) {
            if (dl==dm) {
                il = ilm1;
            } else {
                il = ilp1;
            }
        }
        il += is;
        u[il] = il+lmin;
        if (il==ilm1 || il==ilp1) {
            for (int kb=ji; kb!=jb; kb+=is) {
                il += is;
                u[il] = il+lmin;
            }
        }
    }
}
Automatic registration of PP and PS images using dynamic image warping

Luming Liang & Dave Hale

Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401, USA

![Figure 1](image_url)

**Figure 1.** Registration of PP and PS images. Apparent misfits between manually picked events (red curves) in the PP image (a), and corresponding events in the PS image (b) are largely reduced after the registration (c) using dynamic image warping. The red curves are placed in the same locations in all three images.

**ABSTRACT**

The registration of PP and PS images is a process of correlating corresponding seismic reflections in the two images. We use dynamic image warping to automatically estimate a smooth field of vertical shifts. In warping the PS-image with this shift field, we move reflections in the PS-image to match these in the PP-image. The shift field can also be used to compute the $V_p/V_s$ ratio. Examples for both synthetic and real data show the effectiveness of this method for seismic image registration.

**Key words:** Seismic image registration; converted-wave; $V_p/V_s$ ratio; dynamic image warping

**1 INTRODUCTION**

Registering PP and PS images is an important but difficult task in exploration geophysics. Because of differences between propagation and reflection of P waves and S waves, the same geologic layers appear differently in both positions and waveforms within PP and PS images, as shown in Figure 1a and 1b. The goal of PP and PS image registration is to warp or deform one image to another and thus align them. Useful subsurface information can be extracted from the
estimated warping, e.g., the $V_p/V_s$ ratio (Fomel et al., 2003; Nickel and Sonneland, 2004; Yuan et al., 2008), which can be used in estimating anisotropic parameters (Grechka et al., 2002) or in building correct migration models (Yan and Sava, 2010).

This registration task is not trivial. First, the registration is based on the assumption that corresponding PP and PS images represent the same subsurface features. However, this assumption is not always valid because P and S waves may respond differently to some reflectors. Second, the process may suffer from cycle skipping, especially when shifts change rapidly. To alleviate these difficulties, manual picking of corresponding features is often used.

We propose an automatic image registration technique that can align PP and PS images. This method is based on a global optimization technique called dynamic image warping (DIW) (Hale, 2012). The output of the dynamic image warping is a vertical-shift field, which can then be used to estimate the $V_p/V_s$ ratio.

In this paper, both PP and PS images are post-stack time-migrated images.

2 THEORY

Registration of a PP image $f(x,t_{pp})$ and a PS image $g(x,t_{ps})$ requires a shift field $u(x,t_{pp})$ such that

$$f(x,t_{pp}) \approx g(x,t_{pp} + u(x,t_{pp})),$$

where $t_{pp}$ is the PP two-way travel time and $x$ denotes horizontal CMP distance. Here, we assume the PP and PS images can be aligned with only vertical warping.

From estimates of $u(x,t_{pp})$, we can derive the $V_p/V_s$ ratio. Suppose a reflector located at depth $z$ appears in the PP image at time $t_{pp}$ and appears in the PS image at time $t_{ps}$. Assuming that $t_{pp}$ and $t_{ps}$ are vertical two-way times, we have

$$\frac{dt_{pp}}{dz} = \frac{2}{V_p(z)},$$

$$\frac{dt_{ps}}{dz} = \frac{1}{V_p(z)} + \frac{1}{V_s(z)},$$

where $V_p(z)$ and $V_s(z)$ are P- and S-wave velocities. Therefore,

$$\frac{dt_{ps}}{dt_{pp}} = \frac{1 + V_p(z(t_{pp}))/V_s(z(t_{pp}))}{2}$$

and

$$\frac{V_p(z(t_{pp}))}{V_s(z(t_{pp}))} = 2\frac{dt_{ps}}{dt_{pp}} - 1.$$

Let

$$V_p(t_{pp}) \equiv V_p(z(t_{pp})),
V_s(t_{pp}) \equiv V_s(z(t_{pp})),$$

to obtain $V_p/V_s$ ratio as a function of $t_{pp}$ by

$$\frac{V_p(t_{pp})}{V_s(t_{pp})} = 2\frac{dt_{ps}}{dt_{pp}} - 1. \quad (6)$$

Because S-wave velocities are slower than P-wave velocities, we should compress PS images vertically by a certain factor $c$ as the first step in registration with corresponding PP images. For the example shown in Figure 1, $c = 2$. The relationship between PP time and PS time is

$$t_{ps}(t_{pp}) = ct_{pp} + u(t_{pp}). \quad (7)$$

Thus,

$$\frac{dt_{ps}}{dt_{pp}} = c(1 + \frac{du}{dt_{pp}}). \quad (8)$$

Combining equation 8 with equation 6, we obtain

$$\frac{V_p(t_{pp})}{V_s(t_{pp})} = (2c - 1) + 2c\frac{du}{dt_{pp}}. \quad (9)$$

Equation 9 relates the partial derivative of the shift field with respect to PP time to the $V_p/V_s$ ratio. When $c = 1$, this result is consistent with those given by Fomel et al. (2003) and Yuan et al. (2008).

Partial derivatives $\frac{du}{dt_{pp}}$ correspond to time strains (Hale, 2012). The relationship between the $V_p/V_s$ ratio and the partial derivative $\frac{du}{dt_{pp}}$ provides a clue about how to set strain limits when using DIW to register PP and PS images. Suppose $c = 2$; if we set the strain limit as 10%, the estimated $V_p/V_s$ ratio should be within the range of [2.6, 3.4].

3 1D SYNTHETIC EXAMPLE

Synthetic 1D PP and PS seismograms are shown in Figure 2a. These seismograms were generated by accumulating Ricker wavelets for some random reflection coefficients.

We denote PP and PS seismograms as $f(t)$ and $g(t)$, respectively, both of which are functions of time:

$$f(t) = \sum_i r_i w(t - t_{pp}(z_i)),$$

$$g(t) = \sum_i r_i w(t - t_{ps}(z_i)), \quad (10)$$

where $r_i$ is the random reflection coefficient at depth $z_i$, $t_{pp}(z_i)$ and $t_{ps}(z_i)$ are the PP and PS reflection times, and $w(t)$ is the Ricker wavelet. We compute the PP and PS reflection times as follows:

$$t_{pp}(z_0) = 0,$$

$$t_{ps}(z_0) = 0,$$

$$t_{pp}(z_{i+1}) = t_{pp}(z_i) + 2\delta z/V_p(z_i),$$

$$t_{ps}(z_{i+1}) = t_{ps}(z_i) + \delta z/V_p(z_i) + \delta z/V_s(z_i),$$

where $\delta z$ is the depth sampling interval, $V_p(z_i)$ and $V_s(z_i)$ are specified P- and S-wave velocity functions.
Automatic registration of PP and PS images using dynamic image warping

Figure 2. Synthetic PP (red) and compressed PS seismogram (blue) before (a) and after (b) warping the PS seismogram with the estimated shift shown in Figure 4. The PS seismogram in (a) is compressed by a factor of $c = 2$. Differences (c) between PP and compressed PS seismograms before registration (red) are significantly reduced after registration (blue).

Figure 3. Synthetic P-wave (red) and S-wave (blue) velocities.

shown in Figure 3. After we obtain $f(t)$ and $g(t)$, we compress the PS synthetic seismogram by a factor $c = 2$, so that both synthetic seismograms have the same number of samples.

Estimating the $V_p/V_s$ ratio as a function of depth involves three steps:

1. Estimate shifts $u(t_{pp})$ that align $f(t_{pp})$ and $g(t_{pp})$, as shown in Figure 4a.
2. Compute $V_p/V_s$ ratio as a function of PP time $V_p/V_s(t_{pp})$ using Equation 9, as shown in Figure 4b. (Note that we have only 1D shifts here.)
3. Interpolate $V_p/V_s(t_{pp})$ to get the $V_p/V_s$ ratio as a function of depth $V_p/V_s(z)$, shown as the blue curve in Figure 4c.

In step (1), we use dynamic time warping (Sakoe and Chiba, 1978), the 1D version of dynamic image warping, to estimate the time-variant shifts between two sequences. Dynamic time warping computes a sequence of integer shifts $u[0:N-1]$ between two sequences $f[i]$ and $g[i]$ by solving an optimization problem:

$$u[0:N-1] = \arg\min_{l[0:N-1]} D(l[0:N-1]),$$  \hspace{1cm} (12)\

where

$$D(l[0:N-1]) \equiv \sum_{i=0}^{N-1} (f[i] - g[i + l[i]])^2$$  \hspace{1cm} (13)

subject to

$$|u[i] - u[i-1]| \leq 1, \quad |u[i]| \leq L,$$  \hspace{1cm} (14)

where $N$ is the number of samples in the sequence and $L$ is a specified maximum shift. The constraint imposed on $|u[i] - u[i-1]|$ corresponds to the maximum allowed strain in the estimated shifts. However, $|u[i] - u[i-1]| \leq 1$ is a very loose constraint, it implies that 100% strain is permitted. In practice, we modify this constraint to be $|u[i] - u[i-1]| + |u[i-1] - u[i-2]| + \ldots + |u[i-m+1] - u[i-m]| \leq 1$, for some integer $m$, which approximates a strain limit $S = 1/m$.

The integer shifts $u[0:N-1]$ are then smoothed by a Gaussian filter with half-width equal to image sampling intervals divided by the strain limit to obtain the shifts shown in Figure 4a. Derivatives $\frac{du}{dt_{pp}}$ are computed using a finite-difference approximation.

Figure 2b indicates that, after warping the PS seismogram with the shifts $u(t_{pp})$ that we find by dynamic time warping, the PP and PS seismograms are well aligned. After alignment, differences between
Luming Liang & Dave Hale

Figure 4. Estimated shifts $u(t_{pp})$ (a) between PP and compressed PS seismograms, corresponding $V_p/V_s$ ratio (b) as a function of PP time and true (red) and estimated (blue) $V_p/V_s$ ratios (c) as functions of depth.

Figure 5. True (red) and estimated $V_p/V_s$ ratios with different strain limits: 25% (dotted blue), 50% (dashed blue) and 100% (solid blue).

4 2D REGISTRATION

Dynamic image warping (DIW) is an extension of 1D dynamic time warping (DTW) to higher dimensions. Because we assume that PP and PS images can be aligned with only vertical warping, we can use the DIW presented by Hale (2012) to estimate the vertical shift field from two images. This method is an improvement over tree-sequential dynamic programming (Mottl et al., 2002; Keysers et al., 2007); more details can be found in Section 4.2 of Hale (2012).

The constraints imposed on the 2D vertical shift field $u[i,j]$ that we estimate from images are:

$$|u[i,j] - u[i,j-1]| + |u[i,j-1] - u[i,j-2]| + \ldots + |u[i,j-m_1+1] - u[i,j-m_1]| \leq 1,$$

$$|u[i,j] - u[i-1,j]| + |u[i-1,j] - u[i-2,j]| + \ldots + |u[i-m_2+1,j] - u[i-m_2,j]| \leq 1,$$

$$|u[i,j]| \leq L,$$  (15)

for some integers $m_1$ and $m_2$, which approximates a vertical strain limit $S_1 = 1/m_1$ and a horizontal strain limit $S_2 = 1/m_2$. Here, $u[i,j]$ is a sampled version of $u(x,t_{pp})$ in Section 2. $L$ is again the specified max shift.

To align the PP and PS images shown in Figure 1, we set $L = 15$ samples, $S_1$ and $S_2$ to be 25% and 10%, respectively. The limit of the vertical strain $S_1 = 25\%$ constrains the $V_p/V_s$ ratio to be in the range $[2, 4]$.

The estimated vertical shift field is shown in Figure 6a. From this shift field we obtained the $V_p/V_s$ ratio as a function of the PP travel time using equation 9, as shown in Figure 6d. Then, we performed a time-depth conversion similar to the 1D case shown in the previous section to compute the $V_p/V_s$ ratio as a function of depth.
Figure 6. Image registration for different time-strain limits. Time shifts computed from PP and PS images shown in Figure 1 with strain limits 25% (a), 50% (b) and 100% (c). $V_p/V_s$ ratios (d,e,f) computed from those shifts. Warped PS images (g,h,i). All manually picked events (red curves) are placed in the same locations.
Figure 7. The PP image (a), the PS image registered using dynamic image warping (b) and the PS image registered using shifts derived from velocities $V_p$ and $V_s$ published by Grechka et al. (2002) at locations along the blue line (c).

Figure 8. A comparison between the $V_p/V_s$ ratios at CMP = 10km estimated by our method with strain limits 25% (dotted blue), 50% (dashed blue) and 100% (solid blue) and those cited by Grechka et al. (2002, red).

Registration results with 2 other different vertical strain limits 50% and 100% are also shown in Figure 6. From the warped PS images shown in Figure 6g, h and i, one can see that the registration is generally insensitive to the choice of strain limit. However, some unrealistic deformations appear in Figure 6i (indicated by the yellow oval), suggesting that $S_1 = 100\%$ is an inadequate constraint.

For comparison, Figure 8 includes a $V_p/V_s$ function derived from $V_p$ and $V_s$ functions provided in Figure 13 of Grechka et al. (2002). Those functions were computed in an inversion using prestack PP, PS and checkshot data. They represented their subsurface model with piecewise constant parameters, and this representation accounts for the blockiness in the $V_p/V_s$ curve in Figure 8 that we derived from their results. A visual comparison indicates that the $V_p/V_s$ ratios computed using a vertical strain limit of 50% in dynamic image warping best match the $V_p/V_s$ ratios derived from their results.

We can also use the $V_p/V_s$ ratios from Grechka et al. (2002) with equation 9 to solve for time shifts $u(t_{pp})$, which we can then use to warp the PS image to match the PP image. Figure 7c displays the PS image warped in this way alongside the PP image (Figure 7a) and the warped PS image we obtained using DIW (Figure 7b). While our red horizons (located at the same positions in all three of these images) do not correspond exactly to horizons in the blocky model assumed by Grechka et al. (2006), our image registration is comparable to that implied by their model. The only significant differences occur at late times near the bottom of the PS images, where $V_p$ and $V_s$ velocities were not specified for their model; there we simply extrapolated $V_p$ and $V_s$ velocities to these late times.

5 CONCLUSION

Our use of dynamic image warping to register PP and PS images yields vertical shifts that are constrained to not change rapidly in either vertical or horizontal directions. From derivatives of those shifts we may estimate $V_p/V_s$ ratios.

Our registration method requires that we choose limits for strain, the rates at which shifts may change horizontally and vertically. Although our choices should
depend on image sampling intervals, for stacked seismic images we often observe less variation horizontally, and therefore choose the horizontal strain limit to be lower than the vertical strain limit. In the examples shown in this paper, horizontal and vertical limits of 10% and 50% yielded reasonable registrations.

Our registration for one pair of PP and PS images and $V_p/V_s$ ratios derived from estimated shifts are consistent with those computed by others, who used a more comprehensive and more time-consuming process to invert for $V_p$ and $V_s$, as well as anisotropy coefficients. Whereas that inversion process began with an interpretation and selection of several key horizons, corresponding to boundaries in a layered subsurface model, our image registration and computation of $V_p/V_s$ ratios required only the PP and PS images. In this more comprehensive context of seismic inversion, our registration method could be used to facilitate the initial interpretation and correlation of events in PP and PS images.

ACKNOWLEDGMENTS

We’d like to thank our colleagues for their discussions and feedback on this research. We thank Ilya Tsvankin for providing the images shown in Figure 1. Final thanks to Diane for her help on our writing.

REFERENCES


Automatically tying well logs to seismic data

Andrew Muñoz & Dave Hale
Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401, USA

ABSTRACT
Seismic data are recorded and commonly interpreted in vertical two-way time; well logs, measured in depth, must be tied to seismic using a time-depth curve. However, well ties contain a large amount of uncertainty due to errors in the generation of synthetic seismograms and manual matching of synthetic seismograms to seismic traces. Using dynamic time warping, a fast algorithm that optimally aligns two sequences, we shift and warp a synthetic seismogram to match a seismic trace to produce a well tie and time-depth function.

Key words: seismic well logs tie time depth conversion

1 INTRODUCTION
Making well ties has long been considered an art for geophysical interpreters (White and Simm, 2003). The process involves seismic image processing, wavelet creation, estimation of the time-depth curve, geologic interpretation, and manual corrections. Each step in generating a well tie may require quantitative analysis, but ultimately creating a well tie is a lengthy and interpretive process with potential for significant human error.

We use an algorithm commonly used in speech recognition called dynamic time warping (DTW) (Mueller, 2007) to quantitatively and quickly compute well ties and time-depth functions. Figure 1a shows a synthetic seismogram before and after DTW; visually, the tied synthetic matches the seismic trace. This alignment yields an updated time-depth function shown in Figure 1b.

Integrating well logs and seismic data is crucial when estimating subsurface properties. Many procedures for tying well logs to seismic data exist (e.g., Walden and White (1984); White and Hu (1997); White and Simm (2003); Duchesne and Gaillot (2011); Edgar and van der Baan (2011)). Also, many commercial software packages enable seismic interpreters to generate well ties from well logs and seismic traces. They provide graphical interfaces that allow the user to interactively shift, stretch, and squeeze synthetic seismograms to match time-migrated seismic traces. While the user manually matches synthetics to seismic traces, the software automatically updates a time-depth curve. How-
ever, this process is tedious and relies on the skill and
experience of the user.

Steps common to most well tie processing include:

(i) compute reflectivity from sonic and density logs
(ii) convert reflectivity from depth to time
(iii) convolve an estimated wavelet with reflectivity
(iv) match the synthetic with the seismic trace
(v) update the time-depth curve

This paper describes a similar sequence of steps, but
we use DTW to correct errors associated with rough
estimates of the wavelet and time-depth conversion.

2 SYNTHETIC SEISMOGRAM

The freely available Teapot Dome 3D time-migrated
seismic image (Anderson, 2009) has been rotated,
trimmed, and resampled spatially, as described by Hale
(2010). This data set contains hundreds of well logs that
are also described in Hale (2010). Here we use logs from
only one well.

2.1 Well logs

A well with a velocity log and density log is required
to generate a synthetic seismogram. We use well UWI
490252305400, which has a P-wave velocity log and a
density log. The logs span depths from 0.28 km to 1.85
km. Figure 2 shows the velocity, density, and computed
reflectivity logs. The velocity log \( v(z) \) and the density
log \( \rho(z) \) are uniformly sampled functions of depth for
\( z = z_0, z_0 + \Delta z, \ldots, z_0 + (N_z - 1)\Delta z \). Let

\[
r(z) = \frac{v(z + \Delta z)\rho(z + \Delta z) - v(z)\rho(z)}{v(z + \Delta z)\rho(z + \Delta z) + v(z)\rho(z)}
\]

(1)
denote the acoustic reflection coefficient for normal-incidence P-waves.

This reflectivity is an approximation based on the
acoustic impedance and assumes normal incidence. If
there is a significant amplitude variation with offset in
the seismic data, using elastic impedance to generate
reflectivity would be more accurate (Connolly, 1999).

The reflectivity is sampled with a log interval of
\( \Delta z = 6 \) inches, while the seismic trace is sampled with
\( \Delta t = 4 \) ms (roughly every 4 m, depending on seismic
velocities). We combine a wavelet with the reflectivity
to generate a synthetic seismogram.

2.2 Wavelet estimation

The seismic wavelet is complicated in that it varies with
time, space, and frequency. This makes the generation
of an accurate wavelet a difficult task (Angeleri, 1983).
The wavelet has a significant effect on the synthetic seis-
mogram; even if the logs used to compute reflectivity
were accurately calibrated and corrected for errors, an
inaccurate wavelet would thwart many well tie processes
(Edgar and van der Baan, 2011). Many wavelet approx-
imation methods exist (e.g., Angeleri (1983); Walden
and White (1984); Duchesne and Gaillot (2011); Edgar
and van der Baan (2011)), however, each method is
prone to human error.

Two common methods for wavelet extraction are
statistical and deterministic well-tie methods (Edgar
and van der Baan, 2011). The deterministic well-tie
methods require that a well-tie already exists, while the
statistical method extracts an average wavelet from a
specified window of 3D seismic data (Edgar and van der
Baan, 2011).

A simpler approach is to assume a constant wavelet
with a peak frequency equal to that of the seismic trace.
Figure 3a illustrates a Ricker wavelet with a peak fre-
quency of 35 Hz. This peak frequency is computed from
the amplitude spectrum of the seismic trace nearest to
the well.

Before we combine a wavelet with the reflectivity,
we compute an initial time-depth relationship. Initial
time-depth curves are estimated most commonly using
the P-wave velocity log and check shots, if available
(Edgar and van der Baan, 2011). We follow the com-
mon practice of using the P-wave velocity log \( v(z) \) to
Automatically tying well logs to seismic data

255

compute vertical-two way reflection time:
\[ \tau_z(z) = \tau_0 + 2 \int_{z_0}^{z} \frac{d\xi}{v(\xi)}, \quad (2) \]
where \( \tau_0 \) is the best estimate of vertical two-way time to depth \( z_0 \), the shallowest depth sampled for the velocity log. The subscript \( z \) reminds us that the two-way time \( \tau_z(z) \) is a sampled function of depth \( z \).

Using \( \tau_z(z) \) (and ignoring attenuation, multiple reflections, and other effects), we compute a synthetic seismogram \( g_0(\tau) \) as a simple superposition of seismic wavelets \( w(\tau) \), each delayed by a reflection time \( \tau_z(z) \) and weighted by a corresponding reflectivity \( r(z) \) from equation 1:
\[ g_0(\tau) = \int_{z_0}^{z} r(\xi) w[\tau - \tau_z(\xi)] d\xi, \quad (3) \]
which we sample uniformly at times \( \tau = \tau_0, \tau_0 + \Delta \tau, \ldots, \tau_0 + (N_r - 1)\Delta \tau \). We use \( g_0(\tau) \) to denote the synthetic before amplitude normalization, which is discussed in the next section. Figure 3 shows the wavelet, reflectivity, and synthetic seismogram.

### 2.3 Amplitude normalization

Attenuation in seismic data can cause significant amplitude changes with depth. We therefore use normalization to scale amplitudes in the synthetic seismogram to be similar to amplitudes in the seismic trace. This normalization is time varying. We simply divide each sample by the rms of values nearby.

The amplitude normalization is applied to the seismic trace \( f_0(t) \) and the synthetic seismogram \( g_0(\tau) \), as shown in Figure 4. The normalization has a negligible effect on the seismic trace indicating that the seismic trace is already normalized and any noticeable attenuation was previously removed. However, normalization of the synthetic seismogram has a greater effect on its amplitudes, primarily because the amplitude of the wavelet used to generate the seismogram is incorrect. After normalization, amplitudes in the synthetic \( g(\tau) \) and seismic trace \( f(t) \) are comparable.

### 3 DYNAMIC TIME WARPING

We find a well tie and updated time-depth function using dynamic time warping. Dynamic time warping (DTW) is an algorithm that finds the optimal warping path between two sequences (Mueller, 2007). DTW utilizes dynamic programming to break up an optimization problem into many smaller subproblems, thereby significantly reducing the computational complexity. Thorough explanations of the algorithm are provided in Hale (2012) and Mueller (2007). We follow the three-step process outlined by Mueller (2007).

To align a synthetic seismogram \( g(\tau) \) with a seismic trace \( f(t) \), we use DTW to compute a mapping between \( \tau \) and \( t \) such that
\[ f(t) \approx g(\tau). \quad (4) \]
We use the two artificially generated sequences shown in Figure 5 to demonstrate DTW.

The first step in DTW is to compute an array of squared alignment errors:
\[ e_{i,j} = e(t_i, \tau_j) = [f(t_i) - g(\tau_j)]^2, \quad (5) \]
for all sampled times \( t_i = t_0 + i\Delta t, i = 0, 1, \ldots, N_t - 1 \), and \( \tau_j = \tau_0 + j\Delta \tau, j = 0, 1, \ldots, N_\tau - 1 \). The array of errors \( e_{i,j} \), computed for the sequences in Figure 5, is shown in Figure 6a.

The second step is to recursively accumulate the errors to obtain an array of distances (Hale, 2012). We
The difference between the seismic trace before and after \( f_0(t) \) and \( f(t) \) normalization (a) is negligible. The difference between the synthetic seismogram before and after \( g_0(\tau) \) and \( g(\tau) \) normalization (b) is significant.

**Figure 6.** The array of squared alignment errors (a) for the two sequences in Figure 5. The array of distances (b) is defined by equations 6 and 7. A path with small alignment errors is apparent between sample indices of \( i = 250 \) and \( i = 550 \).

First initialize:
\[
\begin{align*}
    d_{0,0} &= c_{0,0} \\
    d_{0,j} &= c_{0,j} + d_{0,j-1}; & j &= 1, 2, \ldots, N_r - 1 \\
    d_{i,0} &= e_{i,0}; & i &= 1, 2, \ldots, N_r - 1 \\
    d_{i,j} &= e_{i,j} + \min \left\{ \begin{array}{ll}
        d_{i,j-1} & ; j = 1, 2, \ldots, N_r - 1 \\
        d_{i,j-1} & \end{array} \right. \\
    d_{i,1} &= e_{i,1} + \min \left\{ \begin{array}{ll}
        d_{i-1,0} & ; i = 1, 2, \ldots, N_r - 1, \\
        d_{i-1,1} & \end{array} \right. \quad (6)
\end{align*}
\]

which represent distances for samples near the edges of the array (Mueller, 2007). Then, for all \( i > 1 \) and \( j > 1 \), we compute:
\[
\begin{align*}
    d_{i,j} &= c_{i,j} + \min \left\{ \begin{array}{ll}
        d_{i-1,j-1} & \\
        d_{i-1,j-2} + e_{i,j-1} & \\
        d_{i-2,j-1} + e_{i-1,j} & \end{array} \right. \quad (7)
\end{align*}
\]

The resulting array of distances \( d_{i,j} \) is shown in Figure 6b.

The third step is to backtrack in the array of distances \( d_{i,j} \). We denote the top row of the array of distances by total distances \( D_i = d_{i,N_r-1} \), and we set indices \( i^* = \arg \min_i D_i \) and \( j^* = N_r - 1 \) from which we backtrack to obtain a sequence of sample-index pairs \((i^*, j^*)\) (a warping path) that minimizes the distances. From the initial \((i^*, j^*)\), we find the smallest of three values from the right side of equation 7 and extend the warping path with the corresponding sample-index pairs. If the smallest value is \( d_{i-1,j-1} \), the next sample-index pair is \((i^* - 1, j^* - 1)\). If the smallest value is \( d_{i-1,j-2} + e_{i,j-1} \), the next sample-index pairs are \((i^*, j^* - 1)\) and \((i^* - 1, j^* - 2)\). If the smallest value is \( d_{i-2,j-1} + e_{i-1,j} \), the next sample-index pairs are \((i^* - 1, j^*)\) and \((i^* - 2, j^* - 1)\). Notice that we add two new sample-index pairs to the sequence for two of the choices. We then recursively compute the next pair of indices from either \((i^* - 1, j^* - 1)\), \((i^* - 1, j^* - 2)\), or
Automatically tying well logs to seismic data

Figure 7. A schematic of the potential warping paths from \((i^*, j^*)\) in the backtracking procedure. Black dots represent samples in the array of distances, while red arrows are potential additions to the warping path.

\((i^* - 2, j^* - 1)\) (depending on the smallest corresponding value previously found) until either \(i^* = 0\) or \(j^* = 0\).

Figure 7 illustrates the sample-index pairs and potential warping paths in the backtracking procedure. The distance computation and backtracking procedures include constraints described by Sakoe and Chiba (1978). These do not allow warping paths to have long horizontal or vertical segments.

Note that the values of \(D_i\) are sums of squared alignment errors, but that some total distances \(D_i\) are computed from fewer errors than others, potentially making them smaller. Hence, we must normalize by the number of samples that contribute to each \(D_i\) by first backtracking from all samples \((i, N_e - 1)\) for \(i = 1, 2, \ldots, N_e - 1\), and dividing each \(D_i\) by its backtracked path length. This normalization removes the bias of path length and results in a new array of normalized total distances \(\hat{D}_i\).

Figure 8a shows the array of distances with paths backtracked from every sample in \(D_i\). We divide each sample in \(D_i\) by its corresponding path length to obtain \(\hat{D}_i\) shown in Figure 8b. We then backtrack again from the smallest value of \(\hat{D}_i\) to obtain an optimal warping path.

For the artificial example, the backtracked path from the sample corresponding to the smallest value of \(\hat{D}_i\) is consistent with the optimal warping path shown in Figure 8c. The warping path resembles the sinusoidal warping used to generate \(g(\tau)\), and the algorithm correctly computes \(\tau_0^*\), the updated \(\tau_0\). This solution is confirmed by Figure 9, where the warped \(\tilde{g}(\tau)\) is overlaid on the original sequence \(f(t)\), and the two sequences match precisely.

4 TEAPOT DOME DATA

We next apply the DTW algorithm to a Teapot Dome synthetic seismogram \(g(\tau)\) and seismic trace \(f(t)\). We find the arrays of errors and distances shown in Figure 10. The optimal path is not easily seen as it was for the artificial example in Figure 6b. The lack of an obvious optimal path is attributed to the shortcomings in the simple process used to compute the synthetic seismogram \(g(\tau)\).

We use the same backtracking and path length normalization procedure to compute \(\hat{D}_i\) in Figure 10b. We also apply additional constraints to the warping paths as seen in Figure 11c. We assume the entire synthetic seismogram can be tied to the seismic trace, so the optimal warping path must span \(j = 0\) to \(j = N_e - 1\). Also, the synthetic seismogram must be within a window of the seismic trace corresponding to reasonable velocities for the log’s depth. We limit the range of velocities from 1.5 km/s to 7 km/s; the velocity log in Figure 2 validates this range as reasonable. These constraints limit the number of potential warping paths. Notice the large reduction in the number of warping paths from Figure 11a to Figure 11c.
Figure 10. The array of squared alignment errors (a) and distances (b) for the Teapot Dome data. Times $t$ correspond to the seismic trace, and times $\tau$ correspond to the synthetic.

Figure 11. Paths backtracked (a) from every sample in the top row overlaid on the array of distances. Values in the normalized top row of the array of distances $D_i$ (b) do not display a clear minimum. Additional time constraints reduce (c) the number of potential warping paths. The optimal warping path (d) corresponds to the smallest value of $D_i$ and follows valleys of distances. All of these results are for an 84-degree phase rotation of the synthetic seismogram.

Figure 12. The original synthetic $g(\tau)$ (a) is overlaid on a portion of the seismic trace $f(t)$. The warped synthetic $\tilde{g}(\tau)$ (b) has an 84 degree phase rotation. $\tau^*(t)$ (c) is the function used to warp the synthetic seismogram.

To optimally match the synthetic seismogram $g(\tau)$ to the seismic trace $f(t)$, we use $\tau^*(t)$ (Figure 12c):

$$\tilde{g}(t) = g(\tau = \tau^*(t))$$

for a range of times $t \in [t_{\text{min}}, t_{\text{max}}]$ defined by the backtracking step in DTW; $t_{\text{max}}$ and $t_{\text{min}}$ correspond to the first and last sample-index pairs of the backtracked path respectively. In Figure 12b, we show the warped synthetic $\tilde{g}(\tau)$ overlaid on the seismic trace $f(t)$ to illustrate the alignment. Figure 12a shows the original synthetic $g(\tau)$ overlaid on the seismic trace $f(t)$ to illustrate the alignment before warping. The warped synthetic is better aligned than the original synthetic.

4.1 Well tie

From the optimal sequence of sample-index pairs $(i^*, j^*)$ (warping path), we compute two uniformly sampled and monotonically increasing functions $\tau^*(t)$ and $t^*(\tau)$ from each pair of indices where $f(t_{i^*}) \approx g(\tau_{j^*})$. We apply a constant phase rotation to the synthetic seismogram to roughly approximate the seismic trace phase. The optimal path shown in Figure 11d corresponds to the minimum $D_i$ for an 84-degree phase rotation (as described in the next section) in the synthetic seismogram.
4.2 Phase rotation

The phase of the synthetic seismogram should match the phase of the seismic data, but we do not know the seismic wavelet. We approximated the amplitude spectrum by choosing the peak frequency of a Ricker wavelet to match that of the seismic trace. We now approximate the phase with a constant phase rotation that is frequency-independent.

We use DTW to see the effect that phase has on the minimum value of $D_i$ with the goal of further reducing $D_i$. We rotate the phase of the synthetic by 1 degree and measure the minimum $\hat{D}_i$ within the constraints illustrated by Figure 11c. For a 1-degree constant phase rotation, we do not expect the warping path to change drastically, so we apply the constant phase rotation in 1-degree increments for 360 degrees as shown in Figure 13; for each 1-degree increment, we measure the minimum $\hat{D}_i$. At an 84-degree phase rotation $D_i$ is minimized.

4.3 Time-depth function

As stated previously, from the optimal warping path we obtain $t'(\tau)$, and can easily compute a new time-depth function:

$$\tilde{\tau}_s(z) = t'(\tau = \tau_s(z)), \quad (9)$$

where $\tau_s(z)$ is the initial time-depth function computed from equation 2. If $t'(\tau) = \tau$, implying that no warping was necessary, then we obtain $\tilde{\tau}_s(z) = \tau_s(z)$, as expected.

Figure 1b compares $\tilde{\tau}_s(z)$ and $\tau_s(z)$. There is a significant difference between the shape of the two curves and between $\tau_0$ and $\tau'_0$. Future work will further explore the differences of the time-depth functions by computing an interval velocity curve from the updated time-depth function and comparing it to the velocity log.

5 CONCLUSION

Tying wells is tedious and requires multiple processing steps. Many processes exist for tying wells but involve interpretive methods that are prone to human error. Some methods are quantitative, but ultimately, computing accurate well ties is laborious.

We incorporate DTW into a widely used process for tying wells. We use DTW to align a synthetic seismogram and a seismic trace. We apply a constant phase rotation to the synthetic and compute the optimal warping path, which corresponds to minimum $D_i$, the sum of squared alignment errors normalized by the backtracked path length. From the optimal warping path, we compute $\tau'(t)$, from which we warp the synthetic to match the seismic trace, and $t'(\tau)$, from which we compute a new time-depth function $\tilde{\tau}_s(z)$.

Future work will involve computing an interval velocity curve from the updated time-depth function by smoothing the time-depth curve. We also look to further constrain DTW to obtain a more accurate optimal warping path. Furthermore, we will confirm that geologic formations identified from well log measurements match picked seismic horizons of the same formations when the logs are tied to the seismic trace.

DTW gives the seismic interpreter a fast estimate of a well tie. With an increase in accuracy of the synthetic seismogram and DTW constraints, DTW will give a better estimate of a well tied and updated time-depth function.

ACKNOWLEDGMENTS

Thanks to the participants in the C-team seminars who provided many suggestions that furthered this research. Thanks to Diane Witters for her collaborative review of this paper. This work was funded by the Center for Wave Phenomena and its sponsors.

REFERENCES

Duchesne, M., and P. Gaillot, 2011, Did you smooth
your well logs the right way for seismic interpretation?: Journal of Geophysics and Engineering, 8, 514523.
———, 2012, Dynamic warping of seismic images: Center for Wave Phenomena Report 723.
Mueller, M., 2007, Dynamic time warping in information retrieval for music and motion.: Springer. (318 pp.).
ABSTRACT

Earthquakes and active-source seismic surveys provide estimates of depths of subducting slabs, but only at scattered locations. Constructing a useful 3D model of slab geometry involves fitting the depth estimates with a uniformly sampled function of space. The method used to fit the data should account for the curvature of the earth’s surface as well as data uncertainties. In addition to estimates of depths from earthquake locations, focal mechanisms of subduction zone earthquakes also provide estimates of the strikes of the subducting slab on which they occur. We use estimated strike directions and the earth’s curved surface geometry to infer a model for spatial correlation that guides a blended neighbor interpolation of slab depths. The interpolation method is then modified to account for the uncertainties associated with the depth estimates.

Key words: data fitting, interpolation, subducting slab, tensor-guided, uncertainty
1 INTRODUCTION

Accurate knowledge of the geometry of the interface of a subducting slab has important applications in a variety of seismological analyses such as tsunami modeling and propagation (Wang and He, 2008), seismic hazard assessment (e.g. Global Earthquake Model, http://globalquakemodell.org), tectonic modeling and plate reconstruction, and earthquake source inversions (Hayes and Wald, 2009).

One way of acquiring information about the geometry of a subduction interface is through studying the locations and source mechanisms of the earthquakes that occur on or within the subducting slab. However, earthquakes associated with the subduction process are unevenly distributed. The spatial distribution of such earthquakes is dense in seismogenic sections of the subduction zone and sparse in other areas where seismicity rates are low or nonexistent. For example, shallow sections of subducting slabs (above 80 km) are mostly aseismic. In such areas, in the absence of earthquakes, estimates of slab depths derived from active-source seismic or bathymetry surveys can facilitate a subduction geometry model with higher resolution.

Even after combining earthquake data with bathymetry and active-source seismic surveys, slab depths are sampled only at scattered locations. This situation is illustrated in Figure 1a for the subducting Nazca slab beneath western South America. To construct a useful 3D model of the slab interface, we must interpolate these scattered depths to obtain slab depth as a uniformly sampled function of latitude and longitude.

A large variety of interpolation methods are possible. Several previous studies have attempted to model the slab geometry in subduction zones, particularly in deeper regions below the seismogenic zone. For example, Syracuse and Abers (2006), produced hand-drawn contours to match general Wadati-Benioff Zone structure beneath volcanic arcs; others have produced more generalized multi-regional (Bevis and Isacks, 1984) and global (Gudmundsson and Sambridge, 1998) models.

Hayes et al. (2009) and Hayes et al. (2012) produced Slab1.0 which is a 3D subduction geometry compilation of approximately 85% of the subduction zones worldwide. They used estimates of slab strike derived from earthquake source mechanisms, depths of the earthquakes that those mechanisms represent, and a collection of other complimentary datasets to constrain the geometry of the subduction interface. Their method fits the data with a 3D non-planar surface interpolating between a series of 2D sections which sample the slab geometry every 10 km along the strike of the subduction zone. Each 2D section is constructed using Hermite splines and consists of two parts: a polynomial fit of order 2 or 3 to shallow data (depths<80 km) splined with a polynomial fit of order 3 or 4 to intermediate and deep data.

The slab geometry model produced by Hayes et al. (2012) is distinguished from its predecessors because it (1) focuses mainly on the shallowest part of subducting slabs while simultaneously attempting to honor the deeper structure of the slab, (2) filters the earthquake data sets used to include only well-located events with thrust mechanisms associated with the subduction process, and (3) includes additional data sets, such as bathymetry and subduction interface interpretations obtained from active-source seismic profiles.

Although Hayes et al. (2012) use strike data to determine the azimuth of the 2D cross sections they fit with splines, they do not incorporate strike data directly in their polynomial fitting.

In this paper, we use the same filtered subduction-related earthquake and active-source seismic data sets as Hayes et al. (2012), but use a different method to produce a 3D model for the subduction interface in South America. Our tensor-guided method is more direct in that we interpolate all of the scattered depth samples at once to obtain slab depth as a function of longitude and latitude. We use the strike directions and the earth’s curved surface geometry to construct a metric tensor field that guides the interpolation of slab depths.

A metric tensor field provides a measure of distance that need not be Euclidean. For the interpolation shown in Figure 1b, the tensor field depicted by the white ellipses describes geodesic distance measured on the curved surface of the earth. The ellipses are elongated because they account for the curvature of the earth’s surface when projected onto an equi-rectangular longitude-latitude coordinate system.

The tensors used to guide the data fitting, shown as white ellipses with variable elongations and orientations in Figure 1c, account for not only the curvature of the earth’s surface but also the estimates of slab strike directions taken from moment tensor and focal mechanism data. The ellipses are more eccentric in areas where the slab dips more steeply. The orientations of the ellipses are determined using the slab strike directions.

The metric tensor fields shown in Figure 1b and 1c guide interpolation (and fitting) by decreasing distances in directions in which the ellipses are elongated, the directions in which slab depths are more highly correlated. These metric tensor fields therefore provide models of spatial correlation that may yield more accurate models of subducting slab interfaces.

Most interpolation methods require, either explicitly or implicitly, a model of spatial correlation for the data to be interpolated. However, some methods for interpolation do not permit spatially varying models of spatial correlation like those depicted in Figure 1. For example, kriging methods in geostatistics do not easily allow for nonstationary anisotropic correlation models (Boisvert et al., 2009). In particular, covariance functions commonly used in kriging are not guaranteed to remain valid (positive definite) when used with
2 THE METRIC TENSOR FIELD

The methodology described in this paper relies on the blended neighbor interpolation method developed by Hale (2009). As shown in appendix A, this method consists of two steps, where each step requires a metric tensor field \( \mathbf{D}(\mathbf{x}) \) that defines a measure of distance, or equivalently, a model of spatial correlation.

Different tensor fields may be used in different situations. Hale (2009, 2010) shows examples where metric tensor fields are derived from uniformly sampled images. In other situations, where an image is not available, it may be possible to derive the metric tensor fields from other types of secondary data.

In interpolating subducting slab depths, primary data to be interpolated are estimated slab depths; secondary data, from which we derive the tensor field, are estimated slab strikes. The slab depths and strikes are measurements acquired on the curved surface of the earth as a function of longitude and latitude. Therefore, any metric tensor field that we use to guide the interpolation of depths must account for this curvature.

Hale (2011) studies tensor guided interpolation of scattered data on arbitrary non-planar surfaces and provides a general recipe for construction of a tensor field needed for guiding such an interpolation. Here, we follow the same methodology. However, we do not need the same level of generality as provided by Hale (2011). Therefore, in what follows, we will reproduce some of the relevant results presented in Hale (2011) and show the steps of implementing them for addressing the specific problem at hand.

We define the earth’s surface parametrically by a mapping from 2D longitude-latitude coordinates \( \mathbf{u} \) to 3D Cartesian coordinates \( \mathbf{x}(\mathbf{u}) : U \subset \mathbb{R}^2 \to X \subset \mathbb{R}^3 \), and interpolate slab depths in the 2D space \( U \). In this space each location \( \mathbf{u} \) is specified by longitude \( \phi \) and latitude \( \theta \).

The tensor field required to guide a blended neighbor interpolation of depths is constructed in two steps. First, we ignore the curvature of the earth and define a tensor field in an infinitesimal plane tangent to the earth’s surface, where we can assume a flat geometry. Next, we modify this strike tensor field to account for the curvature of the earth.

2.1 Strike tensor field

We base our method for constructing the strike tensor field on the fact that the slab depths should be most highly correlated in the slab strike directions; the spatial correlation of the slab depths is lowest in the slab dip direction (perpendicular to the strike direction). This fact follows intuitively (Figure 2) from the definitions of the strike and dip directions for a dipping plane. Put another way, the dipping structure of the slab results in
an anisotropic model of spatial correlation for the slab depths. This anisotropy is proportional to the dip angle or slope of the slab; the anisotropy is higher for steeper slopes.

The metric tensor $D$ in the eikonal equation A3 is a $2 \times 2$ symmetric and positive definite matrix (Hale, 2009) with two orthonormal eigenvectors $s$ and $d$ and corresponding to positive and real eigenvalues $\lambda_s$ and $\lambda_d$. $D$ can be graphically represented by an ellipse (e.g., see white ellipses in Figure 1) elongated in the direction of the eigenvector corresponding to the maximum eigenvalue and with axes proportional to the square roots of eigenvalues (Hale, 2011). In general, for a tensor field that represents an anisotropic model of correlation, these two eigenvalues are not equal. Here we assume $0 < \lambda_d \leq \lambda_s$.

In the parametric space $\mathbb{R}^2$, the desired strike tensor field $D(u)$ must be designed so that non-Euclidean distances to samples in the strike direction are shorter; such samples therefore get more weight in the interpolation. This design can be achieved by pointing eigenvector $s$ in the slab strike direction and eigenvector $d$ in the slab dip direction (perpendicular to strike), i.e.,

$$s(u) = \begin{bmatrix} \cos \gamma(u) \\ \sin \gamma(u) \end{bmatrix} \quad \text{and} \quad d(u) = \begin{bmatrix} -\sin \gamma(u) \\ \cos \gamma(u) \end{bmatrix},$$

where $\gamma(u)$ represents the estimated strike angle of the slab (the angle between the estimated strike direction and geographic north) at location $u$ on the earth’s surface.

Now we use eigen-decomposition to construct each tensor $D$ as

$$D = \lambda_s s s^T + \lambda_d d d^T,$$

where $s$ and $d$ are the eigenvectors defined in equation 1, and $\lambda_s$ and $\lambda_d$ are their respective eigenvalues. In constructing the strike tensor field for guiding the blended neighbor interpolation of slab depths, what is important is the ratio of the eigenvalues and not their actual sizes; i.e., the aspect ratio of the tensor ellipses is important, not the actual size of their axes. Therefore, we can normalize the eigenvalues and use $\lambda_s(u) = 1$ and some value of $0 < \lambda_d \leq 1$ in equation 2. We let

$$\lambda_d(u) = \frac{1}{1 + \eta \tan^2 \delta(u)},$$

where $\delta(u)$ is the slab dip angle at location $u$ and $\eta$ is a non-negative real parameter.

Note that with $\lambda_s = 1$ the aspect ratio (eccentricity) of the tensor field ellipses is determined by $\lambda_d$. Therefore, using larger values for $\eta$ amounts to increasing the eccentricity of these ellipses and hence the degree of anisotropy in our model for spatial correlation of slab depths at locations where $\tan(\delta)$, the slope of the slab, is nonzero.

As shown in section 4, a value for the parameter $\eta$ can be determined using a 1D line search and cross-validation. At this point, however, we assume that the value for $\eta$ is known and proceed to the next step which is modifying the strike tensor $D$ in equation 2 to define a metric tensor field that accounts for both strike directions and the curvature of the earth.

### 2.2 Accounting for curvature of the earth

The geodesic distance between two points on a non-planar surface is non-Euclidean. This distance can be calculated by solving an eikonal equation written in the parametric space in which the curved surface is defined (Weber et al., 2008).

In interpolation of subducting slab depths, distances are non-Euclidean because of anisotropy in the model for spatial correlation of depths (inferred from estimated strike directions and the dip angles) and also because of the curvature of the earth’s surface. Therefore, to correctly specify the non-Euclidean measure of distance, the desired metric tensor field must account for a combined effect of slab strikes and curvature of the earth.

We approximate the shape of the earth by a sphere. The gradient operator expressed in spherical coordinates is

$$\nabla = \frac{\partial}{\partial r} \hat{r} + \frac{1}{r \cos \theta} \frac{\partial}{\partial \phi} \phi + \frac{1}{r} \frac{\partial}{\partial \theta} \theta. \quad (4)$$

If we denote the radius of the earth by a constant $R$, then on the earth’s surface $r = R$ and equation 4 becomes

$$\nabla = \frac{1}{R \cos \theta} \frac{\partial}{\partial \phi} \phi + \frac{1}{R} \frac{\partial}{\partial \theta} \theta. \quad (5)$$

Note that in this equation, $R \cos \theta$ and $R$ are the respective metric scale factors for longitude $\phi$ and latitude $\theta$ in the spherical coordinate system. Equation 5 can be written in matrix form as

$$\nabla = \begin{bmatrix} \frac{1}{R \cos \theta} & 0 & \frac{1}{R} \frac{\partial}{\partial \phi} \phi \\ 0 & R & \frac{1}{R} \frac{\partial}{\partial \theta} \theta \end{bmatrix}. \quad (6)$$

or equivalently, following the notation of Hale (2011), as

$$\nabla = \nabla_\phi \nabla_\theta, \quad (7)$$

in which $\nabla_\phi$ and $\nabla_\theta$ are defined by

$$\nabla_\phi = \begin{bmatrix} R \cos \theta & 0 \\ 0 & R \end{bmatrix} \quad (8)$$

and

$$\nabla_\theta = \begin{bmatrix} \frac{\partial}{\partial \phi} \phi \\ \frac{1}{R} \frac{\partial}{\partial \theta} \theta \end{bmatrix}. \quad (9)$$

$\nabla_\phi$ denotes the gradient operator in the parametric longitude-latitude space $U \in \mathbb{R}^2$ defined in section 2. Replacing the gradient operator in eikonal equation A3 with the right hand side of equation 7, we can write the
Figure 3. Scattered estimates of subducting slab depths (a) and estimates of uncertainties (b) specified as the standard deviation of the errors associated with depth data. The solid black line represents the west coast of South America. The data points depicted here have been gridded according to the procedure described in the text and consist of 344 measurements from active-source seismic surveys (data points along linear tracks perpendicular to the coastline), and 2057 measurements from bathymetry surveys (data forming a stripe, parallel to the coastline).

This equation is mathematically similar to equation A3 in every aspect except that it is expressed in the longitude-latitude parametric space $U$. We can use tensor field $D_u$ defined in equation 11 to guide the blended neighbor interpolation of slab depths on the non-planar spherical surface of the earth.

### 3 INTERPOLATING SLAB DEPTHS

#### 3.1 Data and initial gridding

Primary data used in this study are scattered estimates of depth of the subducting slab in South America (Figure 3a). Each depth estimate is considered to be a random variable with an associated uncertainty (Figure...
Scattered strike data (a) are interpolated on the curved surface of the earth to construct a uniformly sampled strike field (b). The strike field shown in (c) is the result of the 40th iteration of interpolating the scattered strikes guided along the slab dip direction. White ellipses in (b) represent the tensor field $G$ that was used to guide the interpolation on the curved surface of the earth. These ellipses are elongated as they account for the curvature of the earth’s surface when projected onto an equi-rectangular longitude-latitude coordinate system. The white ellipses in (c) are elongated in the slab dip direction and are used to guide the interpolation of strikes in the dip direction.

Following Hayes and Wald (2009), we assume a normal vertical probability density function (pdf) for depth errors. For earthquake data, the half-width (standard deviation $\sigma$) of the normal pdf is chosen based on the location uncertainty reported in the earthquake catalog from which the data were extracted. For active seismic and bathymetry data, the pdf half-width $\sigma$ is based on an assumed uncertainty related to time-to-depth conversion or estimates of sediment thickness (Hayes et al., 2012).

Secondary data consist of scattered estimates of the strike direction of the slab (Figure 4a). The strike data are taken (Hayes and Wald, 2009) from best-fitting double couples of CMT solutions when available (using the gCMT catalog, www.globalcmt.org).

As a first step, we use a uniform rectangular mesh with grid cells of size $0.1^\circ \times 0.1^\circ$ to grid the depth and strike data separately. The value assigned to each grid cell containing bathymetry, active-source seismic, or strike data is the average of all sample values contained in the cell. No values are assigned to the grid cells that contain no strike or depth data.

For earthquake depth data, we use a weighted averaging scheme described in Appendix B, i.e., the depth value and the uncertainty assigned to a grid cell that contains more than one earthquake sample are computed according to equations B2, B4, and B5. This is because the uncertainties associated with earthquake depths are variable and uncorrelated. By using this weighted averaging scheme, we ensure that more certain depth estimates are given more weight than less certain estimates.

Figure 3a shows the location and values (depicted as colors) of the 5076 depth samples after gridding, and Figure 3b shows the uncertainties associated with these samples on a longitude-latitude map. The location and
values of the 458 gridded strike samples are depicted in Figure 4a. The colors represent the estimates of the strike angle $\gamma$ of the slab measured relative to geographic north.

3.2 Interpolation of the strikes

To perform a blended neighbor interpolation of slab depths, we must construct a metric tensor field using the process described in section 2. This tensor field must be defined at each point on the interpolation grid. As shown by equations 1 and 2, to define the metric tensor field $D$ at each grid point, we need the value of the strike angle $\gamma(u)$ at that point. Therefore, we must first interpolate the scattered strikes.

As shown in Figure 4a, strike data are only available in the region near the coastline. This makes it more difficult to produce accurate estimates of the strike angles for deeper parts of the slab away from the coast.

However, the strikes of the slab are most highly correlated in the dip direction that is perpendicular to the strike direction. This follows from the definition of strike and dip for a dipping plane (Figure 2). Therefore, to obtain more accurate estimates of the strike angles, we guide the interpolation of the scattered strikes in the dip direction using an iterative scheme which consists of three steps:

(i) Interpolate the strikes using tensor-guided interpolation with a metric tensor field that accounts for only the curved geometry of the Earth’s surface. In this case, the non-Euclidean distances are simply geodesic. Therefore, we can use $D = I$ in equation 11 to construct a geometry tensor field $G$ defined as

$$G = F^{-1}F^{-T}. \quad (13)$$

Figure 4b shows the result of the interpolation of scattered slab strikes as a uniformly sampled strike field $\gamma_g(u)$. Tensor field $G$ used to guide this interpolation is represented by the white ellipses.

(ii) Compute a dip direction field $\gamma_d(u)$ using the strike field $\gamma_g(u)$ as

$$\gamma_d(u) = \gamma_g(u) + 90^\circ \quad (14)$$

because the dip direction is perpendicular to the strike direction.

(iii) Construct a tensor field (white ellipses in Figure 4c) based on $\gamma_d(u)$ and use it to guide a new interpolation of scattered strikes, in the dip direction. The result is a new strike field we call $\gamma(u)$.

This is the end of the first iteration. A new iteration starts by going back to step (ii) but this time, instead of $\gamma_g(u)$ in equation 14, we use $\gamma(u)$ obtained in the previous iteration. This process converges (after 40 iterations) to the strike field shown in Figure 4c.

3.3 Interpolation of the depths

With the strike field $\gamma(u)$ obtained in section 3.2, we define tensor field $D$ according to equation 2, where we set $\lambda_1 = 1$ and compute $\lambda_2$ using equation 3 with $\eta = 92$. We chose this value for $\eta$ using the procedure described in section 4.

Next we compute $D_u$, according to equation 11, by multiplying $D$ on the left by $F^{-1}$ and on the right by $F^T$. This is equivalent to combining the strike tensor field $D$ with geometry tensor field $G$ defined in equation 13 to get the combined tensor field $D_u$.

At this point we have all the components needed to perform a blended neighbor interpolation of slab depths on the surface of the earth. Figure 5 shows the results of two blended neighbor interpolations of slab depths guided by two different tensor fields. In Figure 5a, the tensor field $G$ used to guide the interpolation (shown as white ellipses) accounts for only the curvature of the earth’s surface. In Figure 5b, the tensor field $D_u$ used to guide the interpolation (shown as white ellipses) accounts for both that curvature and the strike directions.

4 CROSS-VALIDATION

The solution to any interpolation problem is non-unique and depends on the method used for interpolation; the interpolants shown in Figure 5 are just two solutions among infinitely many. These solutions were both obtained using the blended neighbor interpolation method. However the tensor fields used to guide the interpolations were different.

A possible criterion for assessing the results of an interpolation method is to analyze the accuracy of its predictions. We use a 10-fold cross-validation technique (Kohavi, 1995) to find an optimal value for the parameter $\eta$ in equation 3.

In 10-fold cross-validation, input samples are randomly partitioned into ten mutually exclusive subsets (the folds) of approximately equal size (Kohavi, 1995). Subsequently ten iterations of interpolation and validation are performed such that within each iteration a different subset of samples is held out for validation and the union of the remaining 9 subsets (the training set) is used for interpolation.

The validation process involves computing the numerical difference between interpolated values and test-set sample values normalized by the estimates of uncertainty (standard deviation of error) associated with the samples. Based on this normalized difference, a dimensionless error is assigned to each sample in the test set. Therefore, after 10 iterations, a cross-validation error is assigned to all data samples. The accuracy of the interpolation method can then be assessed by analyzing these cross-validation errors. For instance, we can compute the root mean square (rms) of the normalized
errors and use it as a measure of accuracy of an interpolation method.

Now we explain how to use cross-validation to find an optimal value for $\eta$ in equation 3. The idea is to construct (according to the process described in section 2) different tensor fields using different values for $\eta$, and compute a separate tensor-guided interpolant using each of these tensor fields. We then compute a cross-validation rms normalized error for each of these tensor-guided interpolants. The optimum value for $\eta$ is the one corresponding to the interpolant with minimum cross-validation rms normalized error.

One difficulty with this approach is that equation 3 requires knowledge of dip angles of the slab. To compute the dip angles, we smooth the strike ignorant interpolant of Figure 5 in order to obtain an approximate initial slab model $q_i(u)$. While inaccurate, this initial model is good enough for the purpose of approximating dip angles as

$$\delta(u) \approx \tan^{-1}\|\nabla q_i(u)\|.$$  

(15)
Substituting this approximation into equation 3 yields

$$\lambda_d(\mathbf{u}) = \frac{1}{1 + \eta \|\nabla q_i(\mathbf{u})\|^2},$$  \hspace{1cm} (16)

which is more straightforward to implement than equation 3.

Figure 6 summarizes the results of the cross-validation process used to determine the best choice for parameter $$\eta$$. The blue curve shows the cross-validation rms normalized error as a function of $$\eta$$. The most accurate strike-guided interpolant (the one with the minimum cross-validation rms error) is given by $$\eta = 92$$.

For a quantitative comparison of the strike-ignorant and strike-guided interpolants shown in Figure 5, we contrast their respective cross-validation rms normalized errors inferred from the curve in Figure 6. The rms normalized cross-validation error for strike-guided interpolation with a tensor field constructed using $$\eta = 92$$ is approximately 2.82. The same error for the strike-ignorant interpolation guided by a tensor field constructed using $$\eta = 0$$ is more than three times larger, i.e., 6.34.

5 ACCOUNTING FOR DATA UNCERTAINTIES

The interpolant shown in Figure 5b matches all scattered input depth samples. This is expected as the interpolation error must be (by definition) zero at the location of input samples. However, this interpolant does not represent a geophysically reasonable model for the subduction interface because unlike the interpolant surface of Figure 5b, a model for the slab geometry must be smooth and without abundant inflection points (Hayes and Wald, 2009).

Figure 8a is a vertical cross section showing a profile (green curve) of the interpolated slab depths. Note the abundant fluctuations and excessive curvatures in the green curve. If real, such fluctuations would imply unwarranted bending strains and stresses within the structure of the slab, especially in the shallow parts where the slab is still relatively cold and brittle. Therefore, the fluctuations observed in the interpolated depths are more likely the result of error in depth estimates and do not represent the true geometry of the slab.

To account for data uncertainties and to obtain a smoother model for the slab, we modify the tensor-guided interpolation method to obtain a data fitting method. For this modification, we utilize the given estimates of data uncertainties and incorporate them in a statistically plausible manner into the modeling procedure.

5.1 From interpolation to data fitting

The tensor-guided procedure described so far is based on the blended neighbor interpolation method (summarized in appendix A). The first step of the blended neighbor method can be interpreted as scattering the values $$f_k$$ corresponding to the nearest known sample points $$\mathbf{x}_k$$ to construct a nearest neighbor interpolant $$p(\mathbf{x})$$.

The second step of the blended neighbor method can be interpreted as smoothing $$p(\mathbf{x})$$ to obtain a blended neighbor interpolant $$q(\mathbf{x})$$. The extent of this smoothing at each location $$\mathbf{x}$$ is proportional to the non-Euclidean distance $$t(\mathbf{x})$$ from $$\mathbf{x}$$ to the nearest known sample (Hale, 2009). This means that at the location of known samples, where $$t(\mathbf{x}) = 0$$, no smoothing is applied so the solution to blending equation A4 is $$q(\mathbf{x}) = p(\mathbf{x})$$; hence, the interpolation condition $$q(\mathbf{x}_k) = f_k$$ is satisfied and the interpolant $$q(\mathbf{x})$$ matches all scattered data.

Satisfying the interpolation condition is an implicit requirement of any interpolation method. However, in situations where there is uncertainty associated with data, this condition may not be desirable.

In blended neighbor interpolation, the constraint of satisfying the interpolation condition can be relaxed by adding a function $$w(\mathbf{x})$$ to the term $$t^2(\mathbf{x})$$ in equation A4 so that it becomes

$$q(\mathbf{x}) = \frac{1}{2} \nabla \cdot (t^2(\mathbf{x}) + w(\mathbf{x}))\mathbf{D}(\mathbf{x}) \cdot \nabla q(\mathbf{x}) = p(\mathbf{x}).$$ \hspace{1cm} (17)

This modification implies that the amount of smoothing applied to $$p(\mathbf{x})$$ at every location $$\mathbf{x}$$ (including the locations of known samples $$\mathbf{x}_k$$) can be controlled by choosing a proper function $$w(\mathbf{x})$$. If $$w(\mathbf{x}_k) \neq 0$$ then the solution $$q(\mathbf{x})$$ of the equation above no longer matches all the scattered samples. In this case, $$q(\mathbf{x})$$ does not interpolate but, rather, fits the scattered data. Therefore,
instead of interpolation, we shall refer to the modified scheme described above as tensor-guided fitting.

Different fitting functions can be obtained using different smoothing functions \( w(x) \) in equation 19. In practice, we may require a fitting solution that models some real phenomenon. Therefore it is important to choose a \( w(x) \) that results in the desired fitting solution.

For example, in modeling the subducting slab geometry, the uncertainties are not the same for different types of data and different locations (Figure 3b). The estimates of data uncertainties (variance of the errors) are relatively small for shallow sections of the slab (bathymetry and active-source seismic data) and larger for deeper sections of the slab (earthquake data).

Therefore, the smoothing function \( w(x) \) must be designed so that the amount of smoothing applied to the location of each data sample is proportional to specified data uncertainties. One such smoothing function can be defined as

\[
w(x) = s \sigma^2(x), \tag{18}\]

where \( s \) is a positive scalar parameter and \( \sigma(x) \) is a smooth model of standard deviation that approximates the actual standard deviation of the error associated with data. This model of standard deviation is shown in Figure 7.

Using \( w(x) \) defined above in equation 17, we obtain

\[
q(x) - \frac{1}{2} \nabla \cdot \left(t^2(x) + s \sigma^2(x)\right)D(x) \cdot \nabla q(x) = p(x). \tag{19}\]

The smoothing parameter \( s \) in equation 19 controls the smoothness of the fitting solution \( q(x) \). Note that tensor-guided interpolation is a special case (with \( s = 0 \)) of the more general tensor-guided fitting (with \( s > 0 \)).

The question left to be answered is, therefore, how to choose the smoothing parameter \( s \).

5.2 Choosing the smoothing parameter

In real problems, there are often constraints that make one fitting solution preferable to other possible solutions. For instance, in constraining the geometry of a subduction slab, a geophysically reasonable model for the slab is expected to be smooth. Nevertheless, such a model must not be so smooth that sampled depths are completely disregarded.

Therefore, to obtain an optimum fitting function, we must find a balance between the smoothness of \( q(x) \) and the degree to which it honors the information given by data. Before explaining how to find this balance, we define some new terms.

We define the standardized data error (error in each depth sample value) as

\[
\hat{e}_k = \frac{\mu_k - f_k}{\sigma_k} \tag{20}\]

where \( k \) is the sample index, \( \mu_k \) is the expected value of slab depth (true depth) at location \( x_k \) of the sample, \( f_k \) is the sample value, and \( \sigma_k \) is the uncertainty or the standard deviation of the error associated with the sample. Note that in equation 20, \( \mu_k \) and \( \sigma_k \) are unknown quantities.

Recall that each depth sample is assumed to be a random variable with a normal distribution \( N(\mu_k, \sigma_k) \). Therefore, the standardized data error \( \hat{e}_k \) is expected to be a random variable with standard normal distribution. From this, we infer that the collection of all standardized data errors \( \hat{E} \), computed according to equation 20, also constitutes a population that is a standard normal distribution, i.e.,

\[
\hat{E} \sim N(0, 1). \tag{21}\]

Similarly, we define the standardized fitting error
(residual) at location $x_k$ of each sample as

$$\hat{r}_k = \frac{q(x_k) - f_k}{\sigma_k}$$  \hspace{1cm} (22)$$

where $k$ is the sample index, $q(x_k)$ is the value of the fitting function at location $x_k$ of the sample, $f_k$ is the sample value, and $\sigma_k$ is the uncertainty or standard deviation of the error associated with the sample.

Our goal is to find a fitting function $q(x)$ that correctly estimates the true slab depths. If this goal is attained for some optimum fitting function $q_{opt}(x)$, then we have $q_{opt}(x_k) = \mu_k$ and hence, by equations 20 and 22, $\hat{r}_k = \hat{c}_k$. This implies that, for the optimum solution, the collection of all standardized fitting errors $\hat{R}$ will have a standard normal distribution, i.e.,

$$\hat{R}|_{q_{opt}(x)} \sim N(0, 1).$$  \hspace{1cm} (23)$$

This can be used as a criterion to choose the optimum smoothing parameter $s$ in our data fitting method. In other words, we can analyze fitting errors associated with fitting solutions computed using different smoothing parameters, and then choose the one for which $\hat{R} \sim N(0, 1)$ as the optimal smoothing parameter $s$.

Note that our assumption about the normality of the distribution of the fitting errors might not be accurate. Therefore, when assessing the distribution of the fitting errors, it is important to use a robust statistic which is not severely affected by the potential outliers. One such statistic is the interquartile range (IQR). For $N(0, 1)$, the IQR is the range of values from -0.674 to 0.674 containing 50% of the population. Thus, for an optimum fitting solution, half of the standardized fitting errors are expected to lie within the range $[-0.674, 0.674]$.

To find the right smoothing parameter, we start from the interpolation solution (i.e., from $s = 0$) and then gradually increase the smoothness parameter until we reach a point where half of the fitting errors fall within and half fall outside of the IQR for a standardized normal distribution. We choose the smoothing parameter $s$ for which this condition is satisfied to be the optimal fitting parameter and use it to compute the optimum fitting solution.

6 Results and Discussion

The final result of applying our tensor-guided fitting procedure to slab data is shown in Figure 8c. Compared to the interpolating model (shown in Figure 8b), this fitting model is smoother and is therefore likely to be more geophysically reasonable. The smoothing parameter $s = 0.56$, employed in the tensor-guided fitting to produce the model in Figure 8c was chosen using the method described in section 5.2.

The fitting solution obtained by Hayes et al. (2012) in the Slab1.0 subduction slab compilation is shown in Figure 8d. A cross-section profile of Slab1.0 is contrasted with the tensor-guided fitting and interpolating profiles in Figure 8a. In this cross section, the Slab1.0, the tensor-guided fitting, and the tensor-guided interpolating profiles are the blue, red, and green curves, respectively. Compared to the tensor-guided interpolation, the tensor-guided fitting profile is clearly smoother.

However, our fitting profile (red) is not as smooth as the is Slab1.0 profile (blue). This is because Hayes et al. (2012) use polynomial splines of degrees 2, 3, or 4 to fit the data in 2D sections. Therefore, Slab1.0 solution is forced to be smoother than the tensor-guided fitting result in the cross section shown here. Nevertheless, our model honors the data more precisely than the Slab1.0 model.

Also, note that the 3D slab surface produced by our fitting method (Figure 8c) is smoother than the Slab1.0 model (Figure 8d) along the direction parallel to the coastline. One reason for this is that unlike the method used by Hayes et al. (2012) which requires interpolating between interpolated 2D profiles along the coastline, tensor-guided fitting is performed in one step. Another reason is that the tensor-guided fitting solution is guided along the strike directions of the slab.

In most areas, the slab depths predicted by the tensor-guided method are 0 – 40 km deeper than the depths predicted by the Slab1.0 model. The tensor-guided fitting and interpolating solutions change concavity with their slopes approaching horizontal at the eastern edge of the model (see Figure 8a around distance 900-1070 km and depth 400-600 km). This is due to the use of a zero-slope boundary condition in solving the system of partial differential equations 19.

Figure 9 shows the histogram of standardized fitting errors $\hat{r}_k$ associated with the optimum fitting solution computed using smoothing parameter $s = 0.56$. Note that this histogram does not exactly represent a normal distribution. The long tails of the distribution observed here are not expected for a normal distribution. Therefore, this study might benefit from assuming a different type of distribution (e.g., exponential) for the data and fitting errors.

Also, note that the histogram shown in Figure 9 is not centered at 0, which indicates a bias towards positive fitting errors. As explained below, the observed bias is a consequence of the geometry of the problem, and the smoothing applied in our fitting procedure.

The slab geometry is almost horizontal in areas near the coastline. Therefore, the tensors used in the tensor-guided fitting procedure in these areas have low eccentricity (e.g., see the nearly circular ellipses to the left of the coastline in Figure 5). One consequence of using these tensors in the smoothing step is that the fitting solution $q(x)$ in the bathymetry area receives contribution from the deeper sections of the slab. Thus, $q(x_k)$ is more likely to overestimate the slab depth at locations $x_k$ of the bathymetry samples. This means that the stan-
Figure 8. A cross section (a) showing the profiles of three different slab models. The slab model obtained by tensor-guided interpolation (b) is compared with the model obtained by tensor-guided fitting (c) and the same model from Slab1.0 (d) produced by Hayes et al. (2012). Line segment AB shows the geographical location of the vertical cross section shown in (a). The gray crosses in (a) are the orthogonal projection of all data points (the points shown in white in (b), (c), and (d)) that lie within a rectangular window of width 100 km centered on the vertical plane of section AB. The gray dots in (b), (c), and (d) denote the location of scattered data points.
7 CONCLUSION

Additional information provided by secondary data can be used to guide the interpolation or fitting of a primary set of spatially scattered data. This is done through construction of a metric tensor field that defines a model for spatial correlation. The tensor-guided fitting method presented here is capable of utilizing tensors that are both anisotropic and spatially variable.

We constructed a tensor field based on the earth's curved surface geometry and strike estimates of the subducting slab in South America and used it to guide the interpolation of the scattered estimates of slab depths.

Proper handling of data uncertainties is an important aspect of data fitting methods. If estimates of uncertainties associated with data are available, they can be easily incorporated into the tensor-guided fitting procedure to produce a statistically consistent fit to the scattered data. We used the slab data to demonstrate the capability of our data fitting method in handling data uncertainties.

In other areas of geoscience, such as exploration geophysics, atmospheric, oceanic, and environmental studies, there are problems and applications similar to the one considered in this paper, where the spatial correlation model for one scattered dataset can be inferred from another and where estimates of uncertainty associated with data are available. Therefore, the process of constructing a metric tensor field and accounting for the uncertainties described here can serve as an example for applying tensor-guided fitting in such problems.

8 ACKNOWLEDGMENTS

First author wishes to thank Simon Luo for the helpful discussions on this paper and Diane Witters for her instructions on revision and editing. This research was supported by the sponsors of the Center for Wave Phenomena at the Colorado School of Mines.

APPENDIX A: BLENDED NEIGHBOR INTERPOLATION

Blended neighbor interpolation is specifically designed to facilitate tensor-guided interpolation of scattered data (Hale, 2010).

If we assume the scattered data to be interpolated are a set

$$\mathcal{F} = \{f_1, f_2, \ldots, f_K\}$$

of $K$ known sample values $f_k \in \mathbb{R}$ corresponding to a set

$$\mathcal{X} = \{x_1, x_2, \ldots, x_K\}$$

of $K$ known sample points (coordinates) $x_k \in \mathbb{R}^n$, then the goal of interpolation is to use the known samples to construct a function $q(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ such that $q(x_k) = f_k$.

In the blended neighbor interpolation method (Hale, 2009), the interpolant $q(\mathbf{x})$ is constructed in two steps:

(i) Solve the eikonal equation

$$\nabla t(\mathbf{x}) \cdot \mathbf{D}(\mathbf{x}) \cdot \nabla t(\mathbf{x}) = 1, \quad \mathbf{x} \notin \mathcal{X};$$

$$t(x_k) = 0, \quad x_k \in \mathcal{X}$$

for $t(\mathbf{x})$: non-Euclidean distance from $\mathbf{x}$ to the nearest known sample $x_k$, and

$p(\mathbf{x})$: the value $f_k$ of the sample at point $x_k$ nearest to the point $\mathbf{x}$.

(ii) Solve the blending equation

$$q(\mathbf{x}) = \frac{1}{2} \nabla \cdot t^2(\mathbf{x})\mathbf{D}(\mathbf{x}) \cdot \nabla q(\mathbf{x}) = p(\mathbf{x})$$

for the blended neighbor interpolant $q(\mathbf{x})$.

In the equations above, $\mathbf{D}$ is a metric tensor field which defines the measure of distance in space by providing the anisotropic and spatially varying coefficients of the eikonal equation. In $n$ dimensions, the metric tensor field $\mathbf{D}$ is a symmetric and positive definite $n \times n$ matrix (Hale, 2009).
APPENDIX B: COMBINING MEASUREMENTS HAVING RANDOM UNCORRELATED ERRORS AND KNOWN VARIANCES

Here, we discuss a way to combine independent measurements of a quantity using a weighted averaging scheme.

Consider $N$ independent measurements $(x_1, x_2, \ldots, x_N)$ of the same quantity where each measurement has an unknown expected value and error, i.e.,

$$x_i = \mu + \epsilon_i, \quad i = 1, 2, \ldots, N$$  \hspace{1cm} (B1)

where $\mu$ is the true value of the quantity we wish to estimate, and $\epsilon_i$ is the error in each measurement.

We assume that the errors are not correlated, their expected values are zero, and their variances $\sigma_i^2$ are known.

We let the combined measurement $x$ to be a weighted average (linear combination) of the individual measurements $x_i$,

$$x = \sum_{i=1}^{N} w_ix_i,$$  \hspace{1cm} (B2)

requiring that the weights $w_i$ must satisfy the unbiasedness condition

$$\sum_{i=1}^{N} w_i = 1.$$  \hspace{1cm} (B3)

Using these assumptions, the variance of the combined measurement can be expressed as

$$\sigma^2 = \sum_{i=1}^{N} w_i^2 \sigma_i^2.$$  \hspace{1cm} (B4)

By minimizing this variance with respect to $w_i$ (subject to the unbiasedness constraint of B3) the weights $w_i$ can be determined as

$$w_i = \frac{1}{\sigma_i^2} \sum_{j=1}^{N} \frac{1}{\sigma_j^2}, \quad i = 1, 2, \ldots, N.$$  \hspace{1cm} (B5)

Using these weights, we can compute the best (minimum variance) estimate for the combined measurement $x$ and its variance $\sigma^2$ using equations B2 and B4, respectively. Note that to arrive at these results, we assumed no particular statistical distribution for the measurements.

REFERENCES


Hayes, G. P., and D. J. Wald, 2009, Developing framework to constrain the geometry of the seismic rupture plane on subduction interfaces a priori - a probabilistic approach: Geophys J. Int, 176, 951–964.


Hayes, G. P., D. J. Wald, and K. Keranen, 2009, Advancing techniques to constrain the geometry of the seismic rupture plane on subduction interfaces a priori: Higher-order functional fits: Geochemistry Geophysics Geosystems, 10, 1–19.


Attenuation analysis for heterogeneous transversely isotropic media

Bharath Shekar & Ilya Tsvankin
Center for Wave Phenomena, Colorado School of Mines, Golden CO 80401

ABSTRACT
Attenuation coefficients obtained from seismic data may provide sensitive attributes for reservoir characterization and increase the robustness of AVO (amplitude variation with offset) analysis. Here, we present an algorithm for ray tracing in attenuative anisotropic media based on the methodology of Červený and Pšenčík. Both kinematic and dynamic ray tracing are carried out in an elastic reference medium, with the attenuation terms incorporated as perturbations along the ray. Numerical examples for smoothly varying transversely isotropic (TI) media demonstrate the accuracy of the method.

The dynamic ray-tracing technique provides an efficient forward-modeling operator that can be used in the inversion for anisotropic attenuation. We outline an inversion methodology for estimating attenuation coefficients in 2D heterogeneous anisotropic media. A necessary prerequisite for accurate attenuation analysis is reconstruction of the heterogeneous velocity field. If the subsurface can be approximated by a piecewise-factorized VTI (TI with a vertical axis of symmetry) medium, the anisotropic velocity model can be built using the migration velocity analysis algorithm proposed by Sarkar and Tsvankin. Attenuation coefficients in each factorized block can then be estimated using gradient-based inversion that employs dynamic ray tracing.

Introduction
Seismic attenuation is sensitive to lithology and physical properties of subsurface rocks (Johnston and Toksöz, 1981), so it can provide valuable attributes for reservoir characterization (Lynn, 2004). Compensation for attenuation also helps improve the quality of images obtained by depth migration (e.g., Xin et al., 2008).

The amplitudes of seismic waves are determined by the source characteristics and medium properties including velocity, attenuation coefficients, density, etc. In addition to attenuation, dynamic signatures are influenced by a variety of propagation phenomena including reflection, transmission, mode conversions, triplications, focusing, scattering by heterogeneities, etc. In surface seismic surveys, such factors as source and receiver coupling and near-surface heterogeneities can further distort recorded amplitudes.

Hence, estimation of attenuation from surface seismic data is a challenging problem, and stable attenuation analysis typically requires the subsurface structure to be relatively simple. Brzostowski and McMechan (1992) propose a tomographic algorithm that relies on the spectral-ratio method (Johnston and Toksöz, 1981) to compute isotropic attenuation coefficients from a field data set. They first estimate the 3D velocity structure and then use the simultaneous iterative reconstruction technique (SIRT) (Menke, 1984) to invert for attenuation. While the results are satisfactory for the shallow layers, they deteriorate with depth. The centroid frequency shift method is utilized by Quan and Harris (1997) to construct isotropic attenuation tomograms from a cross-hole data set acquired in West Texas field. Liao and McMechan (1997) apply the same method to a cross-hole survey from Oklahoma to discriminate between sands and shales based on the isotropic attenuation coefficients.

Zhu et al. (2007) extend the spectral-ratio method to anisotropic media and estimate angle-dependent attenuation coefficients from physical-modeling data. Behura and Tsvankin (2009a) introduce a layer-stripping algorithm that combines the velocity-independent layer stripping (VILS) method of Dewangan and Tsvankin (2006) with the spectral-ratio technique to estimate interval attenuation coefficients of pure (PP or SS) reflected waves. The overburden is assumed to be laterally homogeneous (but possibly vertically heterogeneous) with a horizontal symmetry plane, while the target layer may be arbitrarily anisotropic and heterogeneous. By matching time slopes on common-receiver gathers, Behura and Tsvankin (2009a) identify the overburden and target events that share ray segments in the overburden and compute the interval traveltimes and then the inter-
val attenuation coefficient in the target horizon. Reine et al. (2009) propose a similar layer-stripping algorithm to evaluate interval P-wave attenuation in anisotropic media. Since their method operates in the \( \tau - p \) domain, it is restricted to laterally homogeneous target layers.

Shekar and Tsvankin (2011) extend the attenuation layer-stripping method of Behura and Tsvankin (2009a) to mode-converted (PS) waves with the goal of estimating the interval S-wave attenuation coefficient. By identifying PP and PS events with shared ray segments and applying the PP+PS=SS method, they first perform kinematic construction of pure shear (SS) events in the target layer and overburden. Then, the effective shear-wave attenuation coefficients for the overburden and target reflections are obtained from the modified spectral-ratio method. Finally, the dynamic version of VILS is used to compute the S-wave interval attenuation coefficient in the target layer.

The attenuation layer-stripping method, however, is restricted to models with a laterally homogeneous overburden. Here, we build the foundation for extending attenuation analysis to transversely isotropic models with spatially varying velocity and attenuation functions. First, we review the perturbation method of Červený and Pšeničk (2009) for dynamic ray tracing in anisotropic, attenuative media and implement it for heterogeneous TI models. Then we present numerical examples for homogeneous and smoothly varying 2D TI media to verify the accuracy of the algorithm in modeling P-wave attenuation. We conclude by introducing a strategy to invert reflection data for attenuation coefficients in piecewise-factorized VTI media.

**Modeling of seismic wave propagation in attenuative media**

In attenuative media, the stiffness tensor becomes complex, which leads to amplitude decay along seismic rays and velocity dispersion. The time-domain stiffness tensor is called the relaxation tensor, and stress is obtained using the relaxation tensor. Further, to simulate frequency-independent attenuation (“constant-Q” model, e.g., Kjartannson, 1979) it is essential to superimpose various relaxation mechanisms in both isotropic (Xu and McMechan, 1998) and anisotropic (Ruud and Hestholm, 2005) media, thus increasing the cost of finite-difference modeling. Although the approach proposed by Carcione (2011) based on the Fourier pseudospectral method avoids the computation of relaxation functions, it is limited to viscoacoustic media. The reflectivity method (Schmidt and Tango, 1986) can also be used to calculate exact synthetic seismograms in 1D attenuative media. However, it is restricted to laterally homogeneous models with flat interfaces, and not suitable for our purposes.

Therefore, here we compute asymptotic Green’s functions in attenuative anisotropic media using dynamic ray tracing (Červený, 2001). So-called “complex” ray theory treats ray trajectories and parameters computed along the ray as complex quantities (Thomson, 1997; Hanyga and Seredyńska, 2000). However, numerical implementation of “complex” ray theory for seismic applications is not straightforward. Alternatively, ray tracing in attenuative media can be performed using perturbation methods (Gajewski and Pšeničk, 1992; Červený and Pšeničk, 2009), which involves computation of rays in a reference elastic medium with the influence of attenuation modeled as a perturbation along the ray.

Next, we briefly review the ray-tracing methodology of Červený and Pšeničk (2009) which enables the calculation of the asymptotic Green’s function in attenuative, anisotropic, heterogeneous media with smooth spatial variations of the stiffness tensor.

**Dynamic ray tracing in elastic media**

The eikonal equation in elastic, anisotropic, heterogeneous media can be written as:

\[
\mathcal{H}(x_a, p_b) = \frac{1}{2} [G_m(x_a, p_b)],
\]

where \( \mathcal{H} \) is the Hamiltonian and \( G_m \) is the stiffness tensor. No summation is implied over indices \( a \) and \( b \). The term \( G_m(x_a, p_b) \) denotes an eigenvalue of the Christoffel matrix.

Hereafter, we consider in-plane polarized waves in a vertical (\( x-z \)) plane of VTI media, so \( a, b \) and all the other Roman indices take the values 1 and 3. The subscript \( m \) in \( G_m(x_a, p_b) \) will be omitted, and the eigenvalue will be assumed to correspond to P-waves. The Christoffel matrix is given by

\[
\Gamma_{ik}(x_a, p_b) = a_{ijkl}(x_a) p_j p_l,
\]

where \( a_{ijkl} \) is the density-normalized stiffness tensor. Summation over indices \( i = 1, 3 \) and \( l = 1, 3 \) is implied; \( j = i, 3 \) and \( k = 1, 3 \). The eigenvalue of the Christoffel matrix is formally written as

\[
G(x_a, p_b) = g_i a_{ijkl} p_j p_l g_k,
\]

where the components \( g_i \) form the unit polarization vector.

The kinematic ray-tracing equations are given by:

\[
\frac{dx_i}{d\tau} = \frac{\partial \mathcal{H}}{\partial p_i},
\]

\[
\frac{dp_i}{d\tau} = -\frac{\partial \mathcal{H}}{\partial x_i},
\]

with the Hamiltonian \( \mathcal{H} \) defined in equation 1; \( \tau \) represents the traveltime (eikonal) along the ray. Dynamic
Dynamic ray tracing in viscoelastic media

The density-normalized stiffness tensor for attenuative media is complex-valued:

\[ a_{ijkl} = \alpha^{R}_{ijkl} + i \alpha^{I}_{ijkl}, \]

where \( \alpha^{R}_{ijkl} \) and \( \alpha^{I}_{ijkl} \) are the real and imaginary parts of \( \tilde{a} \). The ray-theoretical frequency-domain displacement of P-waves propagating in heterogeneous, attenuative, anisotropic media has the form:

\[ u(x_a, \omega) = S(\omega) A(x_a) \exp\left[-i\omega(t - \tau(x_a))\right] \exp\left[-\omega \text{Im} \tau(x_a)\right], \]

where \( u(x_a, \omega) \) is the displacement vector, \( S(\omega) \) is the source spectrum, \( A(x_a) \) (assumed to be frequency-independent) incorporates the geometrical spreading and transmission coefficients along the raypath, the polarization vector, and the source/receiver directivity. The real and imaginary parts of the traveltime \( \tau(x_a) \) contribute to the phase and amplitude functions along the ray, respectively. The imaginary part of the traveltime is often called the “dissipation factor” and denoted by \( t^* \) (Gajewski and Pšenčík, 1992):

\[ t^*(x_a) = \text{Im} \tau(x_a). \]

The factor responsible for the exponential amplitude decay is called the dissipation filter \( D(\omega) \):

\[ D(\omega) = \exp[-\omega t^*(x_a)]. \]

Červený and Pšenčík (2009) define the perturbation Hamiltonian \( H(x_a, p_b, \alpha) \) as follows:

\[ H(x_a, p_b, \alpha) = H^0(x_a, p_b) + \alpha \Delta H(x_a, p_b), \]

where \( H^0 \) and \( \Delta H \) correspond to the (elastic) reference and (viscoelastic) perturbed medium, respectively, and \( \alpha \) is the perturbation parameter. The Hamiltonian \( H^0 \) can be expressed through the slowness (p) and polarization (g) vectors computed for the reference medium:

\[ H^0(x_a, p_b) = \frac{1}{2} [g_i a_{ijkl} p_j p_k g_k], \]

and \( \Delta H \) is given by

\[ \Delta H(x_a, p_b) = \frac{1}{2} [g_i (a_{ijkl}^R + i a_{ijkl}^I) p_j p_k], \]

where the polarization vector \( \tilde{g} \) computed for the perturbed medium is complex and the asterisk denotes the complex conjugate. The vector \( \tilde{g} \) is found as the eigenvector of the perturbed Christoffel matrix \( \tilde{\Gamma}_{ik} \):

\[ \tilde{\Gamma}_{ik} = (a_{ijkl}^R + i a_{ijkl}^I) p_j p_i, \]

and it satisfies the condition

\[ \tilde{g} \cdot \tilde{g}^* = 1. \]

The traveltime \( \tau(x_a, \alpha) \) and its spatial derivatives \( \partial \tau(x_a, \alpha)/\partial x_i \), which correspond to the perturbation Hamiltonian defined in equation 14, can be expanded in the perturbation parameter \( \alpha \):

\[ \tau(x_a, \alpha) \approx \tau(x_a) + \alpha \left( \frac{\partial \tau(x_a, \alpha)}{\partial \alpha} + \frac{1}{2} \alpha^2 \frac{\partial^2 \tau(x_a, \alpha)}{\partial \alpha^2} \right), \]

(20)

\[ \frac{\partial \tau(x_a, \alpha)}{\partial \alpha} \approx \frac{\partial \tau(x_a)}{\partial x_i} + \alpha \frac{\partial^2 \tau(x_a, \alpha)}{\partial x_i \partial \alpha}, \]

where \( \tau(x_a) \) and \( \partial \tau(x_a)/\partial x_i \) correspond to the reference medium. The partial derivatives in equations 20 and 21 are computed as quadratures along the ray using the reference and perturbation Hamiltonians defined in equations 14–17, and are evaluated for \( \alpha = 0 \).

Equations 20 and 21 can be separated into the real and imaginary part:

\[ \tau(x_a, \alpha) = \text{Re} \tau(x_a, \alpha) + i \text{Im} \tau(x_a, \alpha), \]

(22)

\[ \frac{\partial \tau(x_a, \alpha)}{\partial x_i} = \text{Re} \frac{\partial \tau(x_a, \alpha)}{\partial x_i} + i \text{Im} \frac{\partial \tau(x_a, \alpha)}{\partial x_i}. \]

The dissipation factor and its spatial gradient in the perturbed attenuative medium can be found by substituting \( \alpha = 1 \) into equations 20 and 21 and using equations 12, 22, and 23:

\[ t^*(x_a) = \text{Im} \tau(x_a) \approx \text{Im} \tau, \alpha(x_m) + \frac{1}{2} \tau,\alpha(x_m), \]

(24)
where \( \tau, \alpha, \tau, \alpha = \partial \tau / \partial \alpha, \tau, \alpha = \partial \tau / \partial x_i, \tau, \alpha = \partial^2 \tau / \partial \alpha^2 \) and 
\( \tau, \alpha = \partial^2 \tau / \partial \alpha \partial x_i \).

The P-wave phase attenuation coefficient \( A_P \) is defined as

\[
A_P = \frac{k^I}{k^R},
\]

where \( k^R \) and \( k^I \) are the real and imaginary parts of the wave vector, respectively. The local group attenuation coefficient at any point in space \( x_a \) is given by

\[
A_P(x_a) = -\Im H(x_a) = \frac{1}{2Q(x_a)^2},
\]

where \( Q(x_a) \) is defined as the local quality factor. \( \tilde{Cerven\acute{y}} \) and P\( \tilde{sen\acute{c}ik} \) (2009) prove that equation 27 produces the phase attenuation coefficient computed in the phase direction corresponding to the ray (or to the group angle), and that it is not influenced by the inhomogeneity angle (the angle between the real and imaginary parts of the wave vector). This result agrees with the general perturbation analysis of the influence of the inhomogeneity angle presented by Behura and Tsvankin (2009b).

The exact phase attenuation coefficient in TI media can be found by solving the complex Christoffel equation (\( \tilde{Cerven\acute{y}} \) and P\( \tilde{sen\acute{c}ik} \), 2005; Zhu and Tsvankin, 2006). Zhu and Tsvankin (2006) also present the following linearized approximation for the P-wave phase attenuation coefficient under the assumptions of weak attenuation and weak velocity and attenuation anisotropy:

\[
A_P(\theta) = A_{P0} (1 + \delta_Q \sin^2 \theta \cos^2 \theta + \epsilon_Q \sin^4 \theta),
\]

where \( A_{P0} = 1/(2Q_{P0}) \) is the P-wave symmetry-direction attenuation coefficient and \( \theta \) is the phase angle with the symmetry axis. The attenuation-anisotropy parameters \( \epsilon_Q \) and \( \delta_Q \) depend on the ratios of the real and imaginary parts of the stiffness coefficients and on the velocity-anisotropy parameters (Zhu and Tsvankin, 2006; Tsvankin and Grechka, 2011). We use both the exact and linearized attenuation coefficient in the synthetic examples below.

**Numerical examples**

To test the algorithm described above, we consider an anisotropic halfspace with smoothly varying velocity and attenuation parameters. The velocity parameters \( \epsilon \) and \( \delta \) and attenuation-anisotropy parameters \( \epsilon_Q \) and \( \delta_Q \) are constant, while the P-wave vertical velocity \( V_{P0} \) and vertical quality factor \( Q_{P0} \) vary as linear functions of the spatial coordinates:

\[
V_{P0}(x, z) = V_{P0}^{(0)} (1 + k_x x + k_z z),
\]

\[
Q_{P0}(x, z) = Q_{P0}^{(0)} (1 + j_x x + j_z z),
\]

![Figure 1](image-url) Ray-tracing results for models 1 and 2 from Table 1. (a) Ray trajectories. The quality factor \( Q_P \) as a function of phase angle for (b) model 1 and (c) model 2. The exact coefficient is plotted in gray, the red stars denote the value obtained from equation 27, and the black line is the linearized quality factor obtained using equation 28. (d) Perturbation of the real-valued traveltme as a function of the unperturbed traveltme computed along the ray with a take-off phase angle of 18° in model 2.
where $V_{P0}^{(0)} = V_{P0}(x = 0, z = 0)$ and $Q_{P0}^{(0)} = Q_{P0}(x = 0, z = 0)$. Three sets of the velocity and attenuation parameters used in the tests are listed in Table 1. Models 1 and 2 are homogeneous, while model 3 has linearly varying parameters $V_{P0}$ and $Q_{P0}$ defined by equations 29 and 30. Figure 1a shows a fan of rays with a constant increment in the take-off phase angle from a point source for models 1 and 2. The elastic reference medium for models 1 and 2 is the same, and hence the ray trajectories for the two models coincide. There is a good agree-
Table 1. TI models used to test the algorithm. Models 1 and 2 are homogeneous, while model 3 has linearly varying P-wave vertical velocity $V_{P0}$ and vertical quality factor $Q_{P0}$.

<table>
<thead>
<tr>
<th>Model</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{P0}^{(0)}$ (km/s)</td>
<td>3.00</td>
<td>3.00</td>
<td>3.00</td>
</tr>
<tr>
<td>$V_{S0}^{(0)}$ (km/s)</td>
<td>1.50</td>
<td>1.50</td>
<td>1.50</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>$k_x$ (km$^{-1}$)</td>
<td>0</td>
<td>0</td>
<td>0.20</td>
</tr>
<tr>
<td>$k_z$ (km$^{-1}$)</td>
<td>0</td>
<td>0</td>
<td>0.80</td>
</tr>
<tr>
<td>$Q_{P0}^{(0)}$</td>
<td>20</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>$Q_{S0}^{(0)}$</td>
<td>20</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>$\epsilon_Q$</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>$\delta_Q$</td>
<td>-1.0</td>
<td>-1.0</td>
<td>-1.0</td>
</tr>
<tr>
<td>$j_x$ (km$^{-1}$)</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>$j_z$ (km$^{-1}$)</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 2a displays ray trajectories for model 3, in which $V_{P0}$ and $Q_{P0}$ vary linearly with $x$ and $z$ (Table 1). The perturbation of the real-valued traveltime due to attenuation is almost negligible (Figure 2b). Figures 2c and 2d display the P-wave quality (equation 27) and dissipation (equation 12) factors as a function of the traveltime along the ray. The spatial derivatives of the dissipation factor (Figures 2e and 2f) do not vanish, as expected for this heterogeneous model.

Basic elements of inversion methodology

Due to numerous complications involved in attenuation estimation, robust attenuation analysis typically requires the subsurface structure to be relatively simple. If the overburden is laterally homogeneous and has a horizontal symmetry plane, anisotropic attenuation coefficients of P- and S-waves can be estimated without knowledge of the velocity field using the layer-stripping method of Behura and Tsvankin (2009a) and Shekar and Tsvankin (2011). It should be mentioned that the attenuation-layer stripping method yields attenuation coefficients corresponding to a given source-receiver offset or group angle. To invert for attenuation-anisotropy parameters, it is necessary to estimate the corresponding phase angle, which requires knowledge of an approximate velocity model. However, for more complicated, heterogeneous subsurface models, attenuation estimation requires an accurate velocity field, even if the medium is isotropic.

Velocity estimation in heterogeneous anisotropic media is an active field of research (e.g., Tsvankin and Grechka, 2011). Sarkar and Tsvankin (2004) introduce a 2D method to estimate the P-wave velocity field in VTI media via migration velocity analysis (MVA). They divide the subsurface into factorized VTI blocks, in which the ratios of the stiffness coefficients and, consequently, the parameters $\epsilon$ and $\delta$ are constant. The P-wave vertical velocity in each block is assumed to vary linearly with depth and lateral position (see equation 29). Their MVA algorithm is an iterative two-step procedure that consists of Kirchhoff prestack depth migration followed by velocity updating. In the velocity-analysis step, the residual moveout is estimated in common-image gathers by a 2D semblance scan that takes long-spread (nonhyperbolic) moveout into account. The parameters of each factorized block are updated by minimizing the variance of the migrated depths. The methodology of Sarkar and
Tsvankin (2004) helps resolve the vertical and lateral velocity gradients along with $\epsilon$ and $\delta$, provided that the velocity $V_{P0}$ is known at a single point in each block.

To reconstruct the block-based spatially varying attenuation coefficient, we consider factorized VTI models for both velocity and attenuation (Figure 3). Within each factorized block, the velocity-anisotropy parameters as well as the attenuation-anisotropy parameters are constant, while the P-wave vertical velocity $V_{P0}^{(0)}$ and vertical quality factor $Q_{P0}^{(0)}$ vary linearly with depth and lateral position (equations 29 and 30). The kinematics of P-waves are practically independent of the attenuation coefficient for most subsurface models, except for rocks that exhibit uncommonly strong attenuation (e.g., those saturated with heavy oil, Behura et al., 2007). Therefore, the velocity field can be reconstructed using the algorithm of Sarkar and Tsvankin (2004).

The initial values of the attenuation coefficient for the top layer will be found by applying the spectral-ratio method to pairs of traces of reflections from the layer’s bottom. The layer-stripping method of Behura and Tsvankin (2009a) will then be used to estimate the initial attenuation coefficients in the other layers or blocks. Since the assumptions of that method are not satisfied for models with laterally heterogeneous overburden (Figure 3), it will yield only approximate attenuation coefficients. A gradient-based inversion methodology will then be employed to refine the attenuation estimates. Since the ray-tracing algorithm described above produces the Fréchet derivatives of the complex-valued traveltime along the ray, it can be efficiently used for gradient computation.

Summary

Dynamic ray tracing represents a computationally efficient tool to calculate asymptotic Green’s functions in attenuative anisotropic media. Here, we implemented the algorithm of Červený and Pšeničk (2009), which relies on ray perturbation theory, to model P-wave amplitudes for 2D heterogeneous attenuative VTI media. The attenuation-related terms are computed as perturbations of the reference quantities along rays traced in an elastic background medium. This technique also produces the Fréchet derivatives of the Green’s function, which can be used in solving the inverse problem. We tested the ray tracing methodology on TI models with constant as well as laterally varying velocity and attenuation functions. The algorithm produces accurate P-wave traveltimes and attenuation coefficients, even for models with extremely strong attenuation. The examples also show that the influence of attenuation on traveltime is significant only for media with an uncommonly small quality factor.

Attenuation estimation in heterogeneous TI media has to be preceded by accurate reconstruction of the velocity field. Since attenuation induced traveltime changes are typically negligible, we proposed to perform velocity estimation using the migration velocity analysis (MVA) algorithm of Sarkar and Tsvankin (2004) for piecewise-factorized VTI media. The attenuation coefficients in each factorized block can be represented as linear functions of the spatial coordinates and obtained by gradient-based inversion. The perturbation ray-tracing method substantially reduces the cost of modeling and provides essential quantities for the inversion operator.

Acknowledgments

We are grateful to the members of the A(nisotropy)-Team of the Center for Wave Phenomena (CWP), Colorado School of Mines, for fruitful discussions. Bharath Shekar wishes to acknowledge John Stockwell (CWP) and Sribharath Kainkaryam (Purdue) for their help with ray-theory concepts. Support for this work was provided by the Consortium Project on Seismic Inverse Methods for Complex Structures at CWP and by the Research Partnership to Secure Energy for America (RPSEA).

REFERENCES

Behura, J., and I. Tsvankin, 2009a, Estimation of interval anisotropic attenuation from reflection data: Geophysics, 74, no. 6, A69–A74.
———, 2011, A generalization of the Fourier pseudospectral method: Geophysics, 76.


Thomson, C. J., 1997, Complex rays and wave packets for decaying signals in inhomogeneous, anisotropic and anelastic media: Studia Geophysica etc Geodaetica, 41, 345–381.

Tsvankin, I., and V. Grechka, 2011, Seismology of azimuthally anisotropic media and seismic fracture characterization: SEG.


ABSTRACT
Structural engineers have measured a building’s response to strong motion from civil structures that have been instrumented with accelerometers, such as the Robert A. Millikan Library of the California Institute of Technology. The attenuation of the motion of this building has been measured using seismic interferometry techniques in the past. We use the breaking of the temporal symmetry of the wave equation by attenuation, in combination with seismic interferometry, to estimate attenuation. These estimates are made from fitting the differences in acausal and causal wave forms obtained from different deconvolution processes. We apply the method to the motion recorded at the Millikan Library and obtain estimates of intrinsic attenuation that compare well to past measurements. This technique has more precision for higher frequencies than earlier measurements that are based on seismic interferometry, and it is not dependent on radiation losses at the base of the building.

Key words: seismic interferometry, time reversal, intrinsic attenuation

1 INTRODUCTION
For decades, scientists and engineers have worked to characterize building responses with the purpose of mitigating earthquake hazards and monitoring building integrity (Carder, 1936; Kuroiwa, 1967; Trifunac, 1972; Foutch, 1976; Çelebi et al., 1993; Clinton et al., 2006; Snieder and Şafak, 2006; Chopra and Naeim, 2007; Kohler and Heaton, 2007; Prieto et al., 2010). This has been done by measuring building motion, modal frequencies, intrinsic attenuation, shear velocities, and other properties. After the excitation force drives the motion of the building, intrinsic attenuation, scattering attenuation, and radiation losses dissipate the energy. Intrinsic attenuation estimates quantify the anelastic dissipation of the building’s motion given by the quality factor $Q$ or the damping coefficient $\zeta$.

$$\zeta = \frac{1}{2Q}$$  \hspace{1cm} (1)

Improving the measurement of intrinsic attenuation from the motion excited by complicated ground motion is the focus of our work. Advancing the measurement of attenuation, engineers can more accurately describe the motion of civil structures (Çelebi et al., 1993; Chopra and Naeim, 2007; Kohler et al., 2007), while geophysicists can produce more accurate models of the subsurface (Calvert, 2003) and diagnose the presence of fluids and the migration of these fluids in reservoirs (Bakulin et al., 2007).

Much has been learned from advanced instrumentation installed into buildings such as the Factor Building of UCLA and the Millikan Library of Caltech. These networks have produced large volumes of data for understanding wave propagation in buildings. Monitoring these types of buildings suggests that analysis be done over time to observe changes in the response of the building. Typically this analysis is done for the motion excited by earthquakes (Kohler et al., 2007; Snieder and Şafak, 2006; Clinton et al., 2006), ambient noise (Larose et al., 2006; Derode et al., 2003; Clinton et al., 2006; Prieto et al., 2010), or controlled sources (Kuroiwa, 1967; Clinton et al., 2006; Kohler and Heaton, 2007). Recently, seismic interferometry has been used for acquiring attenuation estimates (Snieder and Şafak, 2006; Kohler et al., 2007; Prieto et al., 2010).

Seismic interferometry has received much attention in the seismology community. The use of seismic interferometry has been explored for extracting Green’s functions, and from this, other parameter estimations (Halliday and Curtis, 2010; Bakulin and Calvert, 2006; Vasconcelos and Snieder, 2008; Wapenaar et al., 2010; Snieder et al., 2006). Deconvolution interferometry is preferable, for reasons discussed later, for retrieving
attenuation estimates. Snieder and Şafak (2006) and Kohler et al. (2007) use deconvolution interferometry for the Millikan Library and Factor Building, respectfully, to acquire attenuation estimates. Our approach is similar, but we use upgoing and downgoing decomposed waves and time reversal to attain attenuation measurements.

The concept that attenuation breaks time reversal symmetry facilitates our measurements of attenuation (Fink, 2006; Gosselet and Singh, 2007). Wavefield decomposition, in combination with deconvolution interferometry, generates acausal and causal waveforms (Snieder et al., 2006). We compare these waveforms, and from their differences, estimate intrinsic attenuation. Since our estimate of attenuation hinges on a comparison of causal and acausal waveforms, we obtain an estimate of intrinsic attenuation because scattering attenuation is causal and invariant for time reversal. In the following we use the term attenuation for intrinsic attenuation.

Much like the method used by Snieder and Şafak (2006), we base our estimates on a linear least-squares fit to the natural logarithm of the deconvolved wave form envelopes. We acquire accurate and precise estimates of attenuation that are frequency dependent and compare well to previous attenuation estimates. Our method acquires frequency-dependent attenuation estimates that do not require any normal mode analysis and estimates are more precise than estimates taken from traveling waves.

The used waveforms are excited by the Yorba Linda earthquake and consist of two N-S and one E-W accelerations of the Millikan Library in 10 floors above the surface and a basement below the surface. Only one N-S acceleration dataset was used for this investigation. There were some instrument coupling issues in the other two datasets as well. Building dimensions and details pertaining to the structure and instrumentation can be found in many articles, but most historical and recently notable, Kuroiwa (1967) and Clinton et al. (2006). Using only the N-S motion of the building constrains our analysis to 1 degree of freedom. The geometry of the building allows for a clamped beam model to represent the motion of the structure. The Millikan Library naturally has 3 degrees of freedom in building motion, and 3n degrees of freedom if we consider n number of floors. Our purpose is to demonstrate the application of this method, but our method can be extended to include more degrees of freedom.

We first discuss the basic theory behind deconvolution interferometry and time reversal. We give an example of why deconvolution interferometry is chosen, and how the deviation from time reversal symmetry indicates the presence of attenuation and facilitates the measurement of attenuation. Next, we discuss the methodology used to achieve these results. We finish by comparing estimates of attenuation with those made from past interferometric methods of the Yorba Linda earthquake data recorded at the Millikan Library.

2 THEORY

Seismic interferometry using deconvolution has become increasingly popular for applications in seismic imaging, parameter estimation, and passive monitoring (Curtis et al., 2006; Snieder and Şafak, 2006; Snieder et al., 2006; Kohler et al., 2007; Vasconcelos and Snieder, 2008; Prieto et al., 2010; Minato et al., 2011). Typically crosscorrelation representation theorems are used to acquire the correct Green’s function between receivers. Snieder (2007) shows that in the presence of dissipation, crosscorrelation type seismic interferometry cannot accurately determine the attenuation response between receivers unless the medium is completely covered by sources. We first briefly explore the reasons why deconvolution seismic interferometry is preferred over crosscorrelation seismic interferometry in estimating attenuation, especially with passive seismic data.

We consider a simple one dimensional seismic interferometry experiment in a homogenous dissipative medium to illustrate our decision to use deconvolution interferometry. Consider a source located at position $r_S$, and receivers located at $r_A$ and $r_B$. (Figure 2). If a dissipating wave propagates away from the source and is recorded by receivers, seismic interferometry can be used to determine the response between those receivers. Seismic interferometry is a tool to measure the response between receivers, where the source position is redatumed to a known receiver location by the virtual source method (Schuster, 2009). Though the source signature of the actual source and virtual source are indeed different, the wave state obtained from seismic interferometry obeys the same wave equation as the original system (Snieder et al., 2006), and we can determine the system response to a virtual source. Examining the de-
Estimating intrinsic attenuation of a building 285

convolution and crosscorrelation operations, for this example, gives insight why deconvolution is preferred for measuring attenuation. In our thought experiment, the receivers record the following frequency-domain wave fields:

\[ U(r_A, \omega) = S(r_S, \omega) e^{-\gamma(r_A - r_S)} e^{ik(r_A - r_S)} \quad (2) \]

\[ U(r_B, \omega) = S(r_S, \omega) e^{-\gamma(r_B - r_S)} e^{ik(r_B - r_S)} \quad , \quad (3) \]

where \( S(r_S, \omega) \) is the source spectrum, \( k \) is the wavenumber, and \( \gamma \) the attenuation coefficient.

The choice of which seismic interferometric operation gives an accurate estimation of the attenuation becomes apparent from the application of each interferometric technique to the wave forms \( U(r_A, \omega) \) and \( U(r_B, \omega) \). Crosscorrelation interferometry applied to these wave fields gives, in the frequency domain,

\[ CC(r_B, r_A, \omega) = U(r_B, \omega) U^*(r_A, \omega) \]

\[ = |S(r_S, \omega)|^2 e^{-\gamma(r_B + r_A - 2r_S)} e^{ik(r_B - r_A)}. \quad (4) \]

The phase is correct, but the amplitude is incorrect because it depends on the sum of the positions \( r_B + r_A \) rather than the difference \( r_B - r_A \). The amplitudes estimated from crosscorrelation is thus incorrect for attenuation analysis. In contrast, deconvolution of the fields of equations (2) and (3) in the frequency domain gives

\[ D(r_B, r_A, \omega) = \frac{U(r_B, \omega)}{U(r_A, \omega)} \]

\[ = e^{-\gamma(r_B - r_A)} e^{ik(r_B - r_A)}. \quad (5) \]

With deconvolution interferometry, we thus obtain the correct phase, \( e^{ik(r_B - r_A)} \), and amplitude, \( e^{-\gamma(r_B - r_A)} \), that accounts for the attenuation of the waves that propagate between the two receivers. Note that crosscorrelation requires the power spectrum of the source signal as well as the source location, as where deconvolution interferometry is independent of the source properties. Because of these properties, we use deconvolved wave forms for our attenuation measurements. Equation 5 is potentially unstable when the reference spectrum \( U(r_A, \omega) \rightarrow 0 \). For our method, we use a stabilized deconvolution given by

\[
D(r_B, r_A, \omega) = \frac{U(r_B, \omega)}{U(r_A, \omega)} \Rightarrow \frac{U(r_B, \omega)U^*(r_A, \omega)}{U(r_A, \omega)U^*(r_A, \omega) + \epsilon}.
\]

where we take \( \epsilon \) to be 1% of the average power of \( U(r_A, \omega) \).

3 METHODOLOGY

Attenuation of waves is expressed in the quality factor that is defined as the relative energy loss over a cycle of oscillation (Aki and Richards, 1980). Estimations of attenuation can be made by measuring the loss in the amplitudes as waves propagate between receivers. These estimations are based on the assumptions that both receivers are coupled accurately to the medium, and the amplitude picks correspond to the same seismic event. Previous studies of Snieder and Şafak (2006) measured attenuation with the Millikan Library data using interferometry to reduce the imprint of a variable receiver coupling. Our method estimates attenuation from individual recordings deconvolved with a common signal, and estimates can be averaged over the array of recordings to reduce and estimate the error. Deconvolution interferometry with a reference signal decomposed into upgoing and downgoing waves, generates wave states with an impulsive upgoing and downgoing wave, respectively, at the base of the building (Snieder et al., 2006). The upgoing and downgoing wave forms of an individual receiver can be compared to give measurements of attenuation.

We separate the wave field at the base of the building into upgoing (\( u_+ \)) and downgoing (\( u_- \)) waves using the following decomposition (Robinson, 1999):

\[
\frac{\partial u_+}{\partial t} = \frac{1}{2} \left( \frac{\partial u}{\partial t} - c \frac{\partial u}{\partial z} \right),
\]

\[
\frac{\partial u_-}{\partial t} = \frac{1}{2} \left( \frac{\partial u}{\partial t} + c \frac{\partial u}{\partial z} \right).
\]
The z-derivative follows from the difference of the motion recorded in the basement and on the first floor. We use the value $c = 322 \text{ m/s}$ as determined by Snieder and Şafak (2006).

Deconvolving the wave forms with their upgoing and downgoing waves separates the signal into causal and acausal waves (Snieder et al., 2006). The causal and acausal wave forms would be symmetric in time if attenuation were not present. We measure the attenuation from the differences in the causal and time-reversed acausal wave forms of each floor. The procedure begins by directionally separating the wave fields in the reference floor using equations (7) and (8). We choose the basement floor recording as our reference signal and separate the signal into upgoing and downgoing waves. We estimate the z-derivative from the differences in the motion recorded in the basement and at the first floor. This generates wave forms that are either causal or acausal in the floors above the basement. This procedure simulates a pure upgoing or downgoing impulsive virtual source in the basement at $t = 0$.

Using only the upgoing waves in the reference signal, our interferometry method extracts the building response from an upgoing impulsive source in the basement of the building at $t = 0$. The deconvolutions of all the floors, in the frequency domain, are ratios of the full wave field spectra of an individual floor and the upgoing wave spectra of the reference floor. We apply this type of deconvolution to all the floors to get eleven deconvolved wave forms (Figure 3a). Deconvolution, with the upgoing waves in the reference floor, compresses all the upgoing waves in the basement into one upgoing virtual impulse injected at $t = 0$. The response of the building to this virtual source is nonzero for times $t > 0$. The use of the upgoing waves for deconvolution generates causal wave forms because the building only oscillates after the upgoing wave enters the building. In a similar manner, we also deconvolve the full wave forms of each individual floor with the downgoing wave form from the basement signal. This downgoing wave is the result of upgoing waves entering the building at earlier times (i.e. $t < 0$). This procedure therefore generates acausal wave forms (Figure 3b). Collapsing all the downgoing waves in the basement to one downward impulse, requires energy to be present in the building before time $t = 0$, which corresponds to the acausal wave forms seen in Figure 3b.

A superposition of the causal and acausal wave forms of Figures 3a and 3b as shown in Figure 3c, reveal a quasi-symmetry of the wave forms around $t = 0$. Time reversal breaks down under certain conditions, such as rotation, flow, and intrinsic attenuation (Fink, 2006). The intrinsic attenuation in the building has broken the time reversal symmetry of the wave forms in Figure 3c. This is apparent from Figure 3d, where the acausal wave forms generated from the deconvolution using downgoing waves of the reference floor, have been time reversed. This plot shows the superposition of the time reversed acausal wave forms and causal wave forms. Note that the amplitudes do not match, and from this difference in amplitudes we measure attenuation.

We band-pass filter the deconvolved signals using Butterworth filters of the 3rd and 2nd order to the respective dominant frequency bands 0.2-3.0 Hz and 5.0-7.8 Hz of the power spectra shown in Figure 4. This allows us to retrieve constant Q values within each frequency band chosen for analysis, and thereby yielding a frequency-dependent Q in a discrete sense. After band-pass filtering our deconvolved signals using the frequency bands of Figure 4, we compute the envelopes of all the wave forms. To demonstrate the estimation of attenuation, we discuss the procedure for the low frequency band-passed data in detail.

The two curves depicted in Figure 5, are a deconvolved wave form from the upgoing wave of floor 4 (gray curve) and the envelope corresponding to this wave form (black curve). This deconvolved wave form is generated from the spectra of the fourth floor and the upgoing
Figure 3. Wave forms deconvolved with decomposed waves and their superpositions both before and after time reversal. a) Wave forms obtained by deconvolving the waves at every floor with the upgoing wave in the basement, b) with the downgoing wave in the basement, c) superposition of causal and acausal wave forms of above figures, d) and superposition after time reversal of acausal wave forms.

wave at the basement. We then apply the low frequency Butterworth band-pass filter of order 3 to this deconvolved wave form. The envelope is acquired from the modulus of the analytic signal using the Hilbert transform. We next take the natural logarithm of this wave form (Figure 6). The second curve is the natural logarithm of the envelope of the wave from the fourth floor obtained from deconvolution using the downgoing wave and time reversed.

The first observation is that these two curves are not the same; this difference is due to attenuation. The second observation is that these curves decay almost linearly for the first 2 seconds. During this time duration, the difference of these natural logarithms is also linear with respect to time, and a constant Q value model corresponds to the slope of the difference of these curves. This model is set up by examining the envelopes of the deconvolved signals of an individual floor. We define the envelopes of the deconvolved signals to be

\[d_+(t) = A_+ e^{-mt}\]
\[d_-(t) = A_- e^{+mt} \]

Taking the natural logarithm of the ratio of \(d_+\) to \(d_-\) yields an equation suited for a linear regression where the slope parameter solves for the attenuation coefficient \(m\) in a least-squares sense.

\[\ln \left| \frac{d_+}{d_-} \right| = -2mt \]

This model does make the assumption that the initial amplitudes are approximately equal, such that \(A_+ \approx A_-\). Figure 7, shows the difference of the curves (black curve) of Figure 6, for the initial 2 seconds, and the linear least-squares fit (gray curve). This procedure is repeated for all the floors, excluding floor 8 because of receiver coupling issues, to generate estimates of the attenuation coefficient.

We also use this procedure for the higher frequency band-pass filtered data of 5.0-7.8 Hz. In this case, the
linear trend of the difference of natural logarithms of the envelopes only has a 1 second duration. The higher frequency content of the signal is expected to lead to a more rapid decay of the envelope than of the low frequency content. Figure 8 shows that noise dominates the signal after 1 second duration because the envelope stabilizes to a near-constant value after that time. Therefore we do our fitting within the first second, and this fitting of the difference of the curves of Figure 8 can be seen in Figure 9.

To estimate the error in our measurement of ζ due to errors in the slope of the fitting curve and the width of the employed frequency band, we use the following equations. If we write the attenuation coefficient as

\[ m = \bar{\omega} \zeta, \]  

where \( \bar{\omega} \) is the weighted mean of the angular frequency for a given frequency band, and the attenuation coefficient \( m \) is given by the slope from our fitting curve, we estimate the error for the \( i \)th floor using

\[ \sigma_{\zeta,i} = \sqrt{\frac{m_i^2}{\bar{\omega}_i^2} \left( \frac{\sigma_{m_i}^2}{m_i^2} + \frac{\sigma_{\bar{\omega}}^2}{\bar{\omega}_i^2} \right)}, \]  

where \( \sigma_\omega \) is our standard deviation of \( \bar{\omega} \) given by

\[ \bar{\omega} = \frac{\int_\Omega \omega P(\omega) d\omega}{\int_\Omega P(\omega) d\omega}, \]  

\[ \sigma_\omega^2 = \frac{\int_\Omega (\omega - \bar{\omega})^2 P(\omega) d\omega}{\int_\Omega P(\omega) d\omega}, \]

where \( P(\omega) \) is the power spectrum within the employed frequency band \( \Omega \). \( \sigma_{m,i} \) is our standard deviation of the attenuation coefficient of the \( i \)th floor from estimates of the discrepancy of the data from the least-squares linear fit (Bevington and Robinson, 2003). This procedure for \( \sigma_{\zeta,i} \) is repeated and averaged for all the floors except the eighth floor. Equation (12) is based on the assumption that the frequency and slope are independent measurements, and therefore that their covariance vanishes. The attenuation estimates with their errors are presented in Table 1.

**RELATION TO PREVIOUS WORK**

The motion of the Millikan Library has been used before to estimate intrinsic attenuation using deconvolution in-
Table 1. Damping Coefficients; TR = time reversal, NM = normal mode, TW = traveling wave

<table>
<thead>
<tr>
<th>Method</th>
<th>ς</th>
<th>σς</th>
<th>Frequency Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>TR</td>
<td>1.14%</td>
<td>0.50%</td>
<td>fundamental</td>
</tr>
<tr>
<td>NM</td>
<td>1.15%</td>
<td>0.41%</td>
<td>fundamental</td>
</tr>
<tr>
<td>Author’s (half-power)</td>
<td>1.57%</td>
<td>NA</td>
<td>fundamental</td>
</tr>
<tr>
<td>Bradford et al.</td>
<td>2.39%</td>
<td>NA</td>
<td>fundamental</td>
</tr>
<tr>
<td>Clinton et al.</td>
<td>1.63%</td>
<td>NA</td>
<td>fundamental</td>
</tr>
<tr>
<td>Kuroiwa (half-power)</td>
<td>1.54%</td>
<td>NA</td>
<td>fundamental</td>
</tr>
<tr>
<td>Kuroiwa (mode shape)</td>
<td>1.47%</td>
<td>NA</td>
<td>fundamental</td>
</tr>
<tr>
<td>Kuroiwa (Hudson’s)</td>
<td>1.74%</td>
<td>NA</td>
<td>fundamental</td>
</tr>
<tr>
<td>TR</td>
<td>1.74%</td>
<td>0.39%</td>
<td>first overtone</td>
</tr>
<tr>
<td>TW</td>
<td>1.58%</td>
<td>1.36%</td>
<td>first overtone</td>
</tr>
</tbody>
</table>

Figure 10. The motion of the building in Figure 1 after deconvolution with the motion recorded in the basement.

Figure 11. Natural log of envelopes from signals in Figure 10 (solid curves), and linear fit (dashed curves)

The technique employing the normal mode oscillations of the building uses the basement floor signal as the reference signal for the deconvolution defined by Equation 6. This technique, however, does not decompose the wave field into up and downgoing waves. Using the full spectra of the reference signal, we generate the deconvolution wave forms in Figure 10. In this figure, the motion at the basement floor is compressed to a band-limited spike at \( t = 0 \). For \( t > 0 \), the figure shows the response of the building to the impulsive excitation. This response is dominated by the fundamental mode of the building with a period of about 0.6 seconds.

We first bandpass filter the waveforms of Figure 10 between 0.2-3.0 Hz, using the low frequency range indicated by the left horizontal bar in Figure 4. A linear curve is fit to the natural logarithm of the envelopes of the deconvolved wave forms similar to our time reversal method. Figure 11 shows the natural logarithm of the envelope of the signals in Figure 10 in solid lines and the least-squares fits in dashed lines. The slope of the least-squares fits is proportional to the attenuation coefficient. Table 1 shows the average damping estimate corresponding to the normal mode measurement indicated NM for this method. An attempt to measure the damping with this method at the higher frequency band yielded poor results. This is due to low amplitude in the power spectrum of the frequency band of 5.0-7.8 Hz.

The next deconvolution interferometry technique developed by Snieder and Şafak (2006) uses higher frequency traveling waves. Snieder and Şafak (2006) make their attenuation measurements using the top floor signal as the reference signal for the deconvolution. Note that there is no decomposition of the wave fields at the reference floor for this traveling wave procedure. The deconvolved wave forms in Figure 12 uses the full spectra of each signal. In Figure 12, a traveling wave moves up and then down the building. Knowing the shear velocity of the building, the ratio of the amplitudes of the upgoing and downgoing waves, the distance traveled from each receiver to the top of the building and back down to the receiver, and the travel times, one can estimate the attenuation. Table 1 displays the results of these measurements marked TW for the traveling wave technique. Since this estimate depends only on the amplitude ratio of the upgoing and downgoing waves at each floor, this estimate is not affected by variations in receiver coupling. The observation that the signals are quiescent after the traveling wave has moved up and down through the building in Figure 12 also suggests there is little scattering caused by the individual floors.

Discussion

Table 1 shows that our method of using deconvolution interferometry with time reversal gives estimates of attenuation that compare well to values found previously using seismic interferometry and classical methods. The new method we propose has several benefits compared to the past methods. Our method recovers attenuation estimates in the normal mode and the first overtone.
This makes the new proposed method more robust than the past interferometric methods because the proposed method’s ability to perform at higher frequencies. Classical methods, such as the half-power method, lose their robustness in higher overtones because of uncertainty in the half-power amplitude picks on either side of the peak frequency for a given overtone.

This method removes the radiation damping from the global attenuation estimate, leaving only internal mechanisms for energy loss like scattering attenuation, constant Coulomb internal friction, and intrinsic material attenuation. This proposed method shares the ability to separate the radiation damping with past seismic interferometric methods. Table 1 gives estimates that employ classical modal analysis that are available in the literature (Bradford et al., 2004; Clinton et al., 2006; Kuroiwa, 1967). These methods recover global damping values of the soil-structure system. These values include radiation damping that occurs at the soil-structure interface. Our method using deconvolution interferometry with time reversal removes the radiation damping by allowing the energy to leave the system. For instance, the upgoing deconvolution used in our proposed method simulates one upgoing wave injected into the building in the basement at time \( t = 0 \). The wave moves up and then down through the building and continues down. In essence there is a reflection coefficient of 0 in the basement. This leaves only internal mechanisms of attenuation in our deconvolved signals from which we measure. Figure 12, of the traveling wave experiment, indicates that there may be little contribution from scattering attenuation. Combining classical methods with our proposed method could refine our knowledge of civil structure behavior and improve earthquake modeling.

Our method is built upon a linear mathematical framework which may lead to limitations in this method and require an attenuation model representation beyond the constant Q description. Our method is described by a linear elastic behavior of the medium, and strong shaking may invalidate the assumption of linearity. Structures proximal to epicenters of large earthquakes will have strong ground motion excitation which can generate nonlinear effects due because the mechanical properties of the building may depend on excitation levels (Clinton et al., 2006). In this dataset for the Millikan Library we did not observe many internal reflections and consequently little scattering attenuation. Datasets with stronger scattering attenuation may introduce complications to this method’s ability to accurately measure intrinsic attenuation.

In the future, this method should be tested on more data to further understand the limitations of this method. Elements of nonlinearity and other sources of attenuation could pose problems for this method. Other civil structures with different geometries such as other building designs, bridges, and downhole arrays could also benefit from this method of attenuation investigation. An extension to include higher degrees of freedom would benefit this method to perform in more complex structures. This new method of deconvolution interferometry with time reversal has proved to be a viable method for this dataset and should be continued to be explored for further advanced applications.

Conclusion

We have shown using data recorded in the Millikan Library that deconvolution interferometry with time reversal is an effective method to measure attenuation in civil structures. This method extracts estimates of intrinsic attenuation from the breaking of time reversal symmetry. By time reversing the acausal wave forms, we estimate intrinsic attenuation using a fitting procedure that is similar to past methods using normal mode oscillations. By comparing the estimates of our time reversal method to that of past seismic interferometry methods, we have shown that our results compare well to the past methods of Snieder and Şafak (2006). Additionally, the time reversal method has higher precision than the traveling wave method and not constrained to measuring only the fundamental mode.

Acknowledgements

This work was supported by the Consortium Project on Seismic Inverse Methods for Complex Structures at Center for Wave Phenomena (CWP). Thank you to the anonymous reviewers for their valuable suggestions that helped improve this article.

REFERENCES

Estimating intrinsic attenuation of a building


Schuster, G. T., 2009, Seismic interferometry: ASA.


3D synthetic aperture for diffusive fields

Allison Knaak\textsuperscript{1}, Roel Snieder\textsuperscript{1}, Yuanzhong Fan\textsuperscript{2}, David Ramirez-Mejia\textsuperscript{2}, \& Mark Rosenquist\textsuperscript{2}

\textsuperscript{1} Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401, USA
\textsuperscript{2} Shell International Exploration \& Production, Houston, TX, USA

\textbf{ABSTRACT}

Controlled-source electromagnetics (CSEM), is a geophysical electromagnetic method used to detect hydrocarbon reservoirs in deep-ocean settings. CSEM is used as a derisking tool by the industry but it is limited by the size of the target, low-spatial resolution, and depth of the reservoir. Synthetic aperture, a technique that increases the size of the source by combining multiple individual sources, has been applied to CSEM fields to increase the detectability of hydrocarbon reservoirs. We apply synthetic aperture to 3D diffusive fields with a 2D source distribution to evaluate the benefits of the technique. We also implement beamforming to change the direction of propagation of the field which allows us to increase the illumination of a specific area of the diffusive field. Traditional visualization techniques for electromagnetic fields, that display amplitude and phase, are useful to understand the strength of the electromagnetic field but they do not show direction. We introduce a new visualization technique utilizing the gradient and phase to view the direction of the diffusive fields. The phase-binned gradient allows a frequency-domain field to appear to propagate through time. Synthetic aperture, beamforming, and phase-binned gradient visualization are all techniques that will increase the amount of information gained from CSEM survey data.

\textbf{Key words:} synthetic aperture, CSEM

\section{1 INTRODUCTION}

Diffusive fields are used in many different areas of science; in this paper we use diffusive fields as an approximation for electromagnetic fields to demonstrate the benefits of synthetic aperture, beamforming, and viewing fields with a visualization technique involving the phase and gradient. This work is motivated by the use of diffusive fields used in controlled-source electromagnetics (CSEM), a geophysical electromagnetic method for detecting hydrocarbon reservoirs in deep-ocean settings. In CSEM, a horizontal antenna is towed just above the seafloor, where seafloor electromagnetic receivers are placed. CSEM was first developed in academia by Charles Cox in the 1970s and since then CSEM has been adopted by the industry and is used for derisking in the exploration of hydrocarbon reservoirs (Constable and Srnka, 2007; Edwards, 2005; Constable, 2010). The electromagnetic field in CSEM is predominantly diffusive because the source is low frequency and the signal propagates in a conducting medium. CSEM has some limitations that keep it from competing with other geophysical methods such as seismic. The size of the hydrocarbon reservoir must be large enough compared to the depth of burial to be detected and the signal that propagates through the reservoir is weak when compared to the rest of the signal (Constable and Srnka, 2007; Fan et al., 2010). Also, CSEM has low spatial resolution compared to seismic methods (Constable, 2010). These drawbacks prompted an investigation into improving the signal received from the hydrocarbon reservoir through synthetic aperture, a method developed for radar and sonar that constructs a larger virtual source by using the interference of fields created by different sources (Barber, 1985; Bellettini and Pinto, 2002). Fan et al. (2010) demonstrated that the wave-based concept of synthetic aperture sources can also be applied to a diffusive field and that it can improve the detectability of reservoirs. The similarities in the frequency-domain expressions of diffusive and wave fields show that a diffusive field at a single
frequency does have a specific direction of propagation (Fan et al., 2010). Once synthetic aperture is applied, the field can be steered using beamforming, a technique used to create a directional transmission from a sensor array (see Haynes, 1998; Van Veen and Buckley, 1988). The basic principles of phase shifts and addition can be applied to a diffusive field to change the direction in which the energy moves. These create constructive and destructive interference between the energy propagating in the field which, with a CSEM field, can increase the illumination of the reservoir (Fan et al., 2012). Manipulating diffusive fields by using interference is not necessarily new; the interference of diffusive fields has been used previously in physics for a variety of applications (Yodh and Chance, 1995; Wang and Mandelis, 1999). Fan et al. (2010) were the first to apply both concepts of synthetic aperture and beamforming to CSEM fields with one line of sources. Fan et al. (2012) demonstrate the numerous advantages of synthetic aperture steering and focusing to CSEM fields with a single line of sources; the detectability for shallow and deep targets greatly improves with the use of synthetic aperture. We extend this work by expanding the technique to a 2D source distribution. In this paper, we introduce the concept of 3D synthetic aperture for diffusive fields, provide examples of steered diffusive fields, present a new visualization technique, and provide examples demonstrating the benefits viewing diffusive fields with a phase-binned gradient.

2 3D SYNTHETIC APERTURE AND BEAMFORMING

2.1 Mathematical Basis

Synthetic aperture was first applied to diffusive fields by Fan et al. (2010) with one line of sources. Before this new use, synthetic aperture was used for radar, sonar, medical imaging and other applications (Barber, 1985; Bellettini and Pinto, 2002; Jensen et al., 2006). One reason synthetic aperture, a wave-based concept, has not been previously applied to diffusive fields is that it was thought diffusive fields could not be steered because they have no direction of propagation (Mandelis, 2000). Fan et al. (2010) showed that the 3D scalar diffusion equation has a plane wave solution at a single frequency with a defined direction of propagation which allows the direction of the field to be manipulated by synthetic aperture. The 3D scalar homogeneous diffusion equation is an appropriate approximation of a CSEM field because at a low frequency and in conductive media, like the subsurface, CSEM fields are diffusive (Constable and Srnka, 2007). In the frequency domain, the 3D scalar diffusion equation in a homogeneous medium, under the Fourier convention $f(t) = \int F(\omega) e^{-i\omega t} d\omega$, is given by

$$D \nabla^2 G(r, s, \omega) + i\omega G(r, s, \omega) = -\delta(r-s), \quad (1)$$

where $D$ is the diffusivity of the medium, $\delta$ is the Dirac-Delta function, $\omega$ is the angular frequency, and $G(r, s, \omega)$ is the Green’s function at receiver position $r$ and source location $s$. For synthetic aperture, we start with a diffusive field created from one point-source. The field from a point-source is given by

$$G(r, s, \omega) = \frac{1}{4\pi D |r-s|^2} e^{-i k |r-s|} e^{-i k |r-s|}, \quad (2)$$

(Mandelis, 2000) where $G(r, s, \omega)$ is the Green’s function at receiver position and source location $s$, $\omega$ is the angular frequency, and $D$ is the diffusion constant. The wave number is given by $k = \sqrt{\omega/(2D)}$. The field from a point-source is the building block for synthetic aperture with diffusive fields. Multiple point-source fields can be summed to create one large source; the interference of the different sources combines to create a synthetic aperture source with greater strength than an individual point-source. The equation for synthetic aperture is given by

$$S_A(r, \omega) = \int_{sources} e^{-A} e^{-i\Delta\Psi} G(r, s, \omega) ds, \quad (3)$$

where, for the source $s$, $\Delta\Psi$ is the phase shift and $A$ is an energy compensation coefficient. A traditional CSEM survey tows a source in parallel lines over receivers placed on the seafloor (Constable and Srnka, 2007); in the following we assume that the sources are towed along parallel lines that are parallel to the x-axis. Then we assume, also, that the y-axis is aligned with the crossline direction. The field is steered by applying a phase shift, for either inline steering or crossline steering, and energy compensation terms defined below:

$$\Delta\Psi = k \hat{n} \Delta s \quad (4)$$

$$\hat{n} = \begin{pmatrix} \cos \varphi \sin \theta \\ \sin \varphi \sin \theta \\ \cos \theta \end{pmatrix} \quad (5)$$

$$A = \Delta\Psi. \quad (6)$$

The phase shift, for an individual source, is shown in equation 4. The shift is a function of the wavenumber, the steering angle $\hat{n}$, and a distance $\Delta s$. Equation 5 defines the steering direction which is controlled by two angles, $\theta$ and $\varphi$. The dip of the direction of the steering angle, represented by $\theta$, is measured with respect to the vertical. The other angle, $\varphi$, represents the azimuthal direction. For inline steering, $\varphi = 0^\circ$ and for crossline steering, $\varphi = 90^\circ$; these are the only directions for $\varphi$ considered in this work because they offer the best steering for a traditional CSEM survey set-up. The quantity, $\Delta s = |s_n - s_i|$ is the distance between an individual source, $s_n$, and the source defined to be at the
The steering angle, \( s_1 \), is defined to be located at the origin. The steering angle, \( \hat{n} \), is a function of \( \theta \), the dip and \( \varphi \) which is either 0° or 90° to steer in the inline or crossline directions, respectively.

In general, the phase shift is defined as \( \Delta \Psi = k(n_x \Delta s_x + n_y \Delta s_y) \). For inline steering, the phase shift equation simplifies to \( \Delta \Psi = k \sin \theta \Delta s_y \) and for crossline steering, the equation simplifies to \( \Delta \Psi = k \sin \theta \Delta s_x \) where \( \Delta s_x \) is the distance between the two sources in the x-direction and \( \Delta s_y \) is the distance between the sources in the y-direction. Figure 1 demonstrates a traditional CSEM survey design with the steering angle and \( s_1 \) labeled. To achieve the final steered field, \( G(r, s, \omega) \) is summed over all sources in each individual line and, then, summed over all lines, as shown in equation 3. The exponential weighting, shown in equation 6, is just one way to create the interference needed to steer diffusive fields (Fan et al., 2012). For a homogeneous medium, the phase shift and energy compensation term are set to be equal because the decay of the field is proportional to the phase shift, and the attenuation coefficient in equation 2 is equal to the wave number (Fan et al., 2011). For a CSEM field, the energy compensation term accounts for the diffusive loss, decreases the background field to create a window to view the secondary field and equalizes the interfering fields to create destructive interference (Fan et al., 2011, 2012). We demonstrate the benefits of synthetic aperture and beamforming with numerical examples.

### 2.2 Numerical Examples

For all of the models shown in the next examples, the model volume is 20km × 20km × 4km to approximate the depth, width, and length of a traditional CSEM survey. We use parallel lines of sources which are the standard survey set-up in industry; in these examples, the 2D source distribution used is constructed from five 5km long tow lines that each contain 50 individual point-sources. The lines are spaced 2km apart which is a common spacing for receivers in the crossline direction. Diffusive fields are difficult to visualize with a linear scale because the field varies over many orders of magnitude. Therefore we use the transformation defined by Fan et al. (2010) to view the field’s amplitude and sign with a logarithm to account for the rapid decay of the field. The transform is shown below:

\[
I_G = m \times \text{Im}(S_A)
\]

\[
Z = sgn(I_G) \log_{10} |I_G|.
\]

The factor \( m \), in equation 7, is a constant scaling factor which sets the smallest amplitude of \( |S_A| \) equal to \( 10^0 \). The dimensionless \( Z \) field displays the log of \( I_G \) with a minus sign when \( I_G \) is negative. The diffusive field from a point-source is shown in Figure 2. All the fields are created at a frequency of 0.25 Hz and a diffusivity \( D = 2.4 \times 10^5 \text{ m}^2/\text{s} \) which is the approximate diffusivity of an electromagnetic wave in seawater (Fan et al., 2010). Figure 2 demonstrates how the field excited by a single point source diffuses through a 3D volume; the strength of the field decreases with increasing depth. We then apply synthetic aperture to the 2D source distribution in the inline and crossline directions. The unsteered field depicted in Figure 3 has five synthetic aperture sources each 5 km long. The sources in the five lines are summed, without any phase shifts and amplitude factors applied, in the x- and y-directions to produce a larger, longer source.

In CSEM the longest source dipole is around 300 meters (Constable, 2010); with synthetic aperture, we can create a much longer source without requiring a boat to tow the extra-long source. Beamforming is applied to the field to change the direction of the energy. For inline steering, individual sources in each of the five lines are multiplied by a phase shift and an energy compensation term and then the sources are summed. The inline direction has more sources to use because the source is towed in the x-direction with samples taken every 100 meters. Figure 4 demonstrates how steering caused the field to be asymmetric towards negative x-values. The diffusive field is steered in the crossline direction much the same way as in the inline direction, but a different phase shift is applied to each synthetic aperture source with all the individual sources on one line multiplied by the same phase shift. As shown in figure 5, this produces an asymmetric movement of the strength of the field in the negative y-direction. It is promising that even with five lines spaced 2 km apart, we can achieve a marked crossline steering of the field; the maximum of the field has been shifted to the right away from the centerline of the survey (\( y = 0 \text{km} \)). This leads us to believe that once applied to CSEM, crossline steering may direct the field toward a target. Inline and crossline steering can be combined to create a field that has energy shifted in the x- and y-directions (Figure 6). The combined steering creates a field that is asymmetric with respect to both axes, concentrating the energy in an area.

Applying synthetic aperture to diffusive fields
Figure 2. The point-source 3D scalar diffusive field log-transformed with the source at (0,0,0).

Figure 3. The unsteered 3D scalar diffusive field log-transformed.

Figure 4. The 3D scalar diffusive field log-transformed and steered at a 45° angle in the inline direction. The arrow shows the shift in the x-direction of the energy from the origin; without steering the maximum energy would be located at the center of the source lines (x=0km).

Figure 5. The 3D scalar diffusive field log-transformed and steered at a 45° angle in the crossline direction. The arrow shows the shift in the energy from the centerline of the survey (y=0km). The shift is less than in Figure 4 because fewer sources are used.
demonstrates the possibilities of using 3D synthetic aperture on real electromagnetic fields acquired from CSEM surveys. The inline and crossline steering of diffusive fields shows how the maximum can be shifted to a new area which allows that area to be illuminated than without steering. The log-transform is a useful tool to view diffusive fields but it has some drawbacks. The only information communicated is the sign and normalized amplitude of the field. There is no information about the direction of the field or a sense of how it propagates in 3D space. We developed a new way to visualize the fields that shows the direction and propagation of the fields in the frequency-domain.

3 3D VISUALIZATION

3.1 Mathematical Basis

The most common way to visualize electric and magnetic field is through magnitude and phase plots but these lack the capability to show the direction the field is traveling (Constable, 2010). Additionally, the log-transformation employed in the previous figures only shows the sign and normalized amplitude of the field, but we need to visualize the direction of the field to identify the enhancement of the upgoing field from synthetic aperture and beamforming because the most important information from a CSEM survey is the electromagnetic signal that propagates down, through the reservoir, and then returns to the seafloor to be recorded by a receiver. This signal is difficult to identify because it is much weaker compared to the background field. The Poynting vector measures the direction in which the energy flux of the electromagnetic field is traveling; it is an effective way to examine how an electromagnetic field propagates (Fan et al., 2012). The energy flux density of the electromagnetic field is given by (Griffiths, 2008)

\[ S \equiv \frac{1}{\mu_0} E \times B, \]  

where \( \mu_0 \) is the permeability of free space, \( E \) is the electric field, and \( B \) is the magnetic field. The diffusive field, an approximation of a diffusive electric field, we use is a scalar field and therefore the Poynting vector cannot be used. The gradient of a scalar field is, however, similar to the Poynting vector for an electromagnetic field: the Poynting vector is the energy flux density of the electromagnetic field which is similar to the heat flux density used in thermodynamics. The heat flux density is found by taking the gradient of temperature which makes the gradient a useful measure of the energy flux for scalar diffusive fields (Schroeder, 1999). The gradient of a diffusive field is complex and only the real part is used to display the gradient, normalized to make the direction apparent. In addition to visualizing the direction of the field, it is useful to know the direction of the field in relation to the time over which the field has propagated; the use of the phase of the field in conjunction with the gradient allows a frequency-domain field to appear to propagate through time. A simple example demonstrates this concept. Consider a point-source field, \( e^{ikr} \) where \( k \) is the wavenumber and \( r \) is the distance from the source to another location in space. In that case, the phase is, then, equal to \( kr \) which smoothly increases with respect to \( r \). A small phase corresponds to a location close to the source and a large phase corresponds to a location farther away from the source. Thus, when the phase is binned by multiples of \( \pi \) the frequency-domain diffusive field appears to propagate outward in space the same way it propagates through time. The phase must not contain phase-jumps for use in our visualization method. To make the phase of the field increase smoothly, we use phase-unwrapping in 3D, which corrects the phase-jumps of \( 2\pi \) that occur in the field. Phase unwrapping is applied in many fields including radar, medical, and geophysics; 3D phase unwrapping is an ongoing field of research (Itoh, 1982; Wang et al., 2011; Parkhurst et al., 2011; Shanker and Zebker, 2010). Phase unwrapping is simple for our noiseless scalar diffusive fields; we unwrap the phase one dimension at a time to construct a smoothly increasing phase. Phase jumps determined to be larger than the tolerance value \( \pi \) are reduced by adding or subtracting \( 2\pi \) to until the jump is less than the tolerance value. Once the phase is unwrapped, we add a constant to the phase field to make the source phase equal to zero. Only the relative
3.2 Numerical Examples

As in the previous section, the volume of the diffusive field is $20\text{km} \times 20\text{km} \times 4\text{km}$ to represent a CSEM survey area. We first demonstrate the phase-binned gradient visualization technique with a simple point-source before applying the method to a field with a synthetic aperture source. The phase and gradient are shown in Figures 7 and 8 to compare with the 3D phase-binned gradient in Figure 9.

The temporal evolution of the field cannot be viewed with the gradient or the phase. The advantage of the phase-binned gradient is that it displays the direction of the energy from the origin outward to the edges of the model as a function of increasing phase. The direction of the field is displayed for different phase bins; a smaller phase corresponds to parts of the field closer to the source and a larger phase corresponds to parts of the field farther away from the source. This type of visualization becomes more useful when examining a field with synthetic aperture and steering which are more complex than the point-source example. Figures 10 and 11 display the unwrapped phase and gradient, respectively, of a scalar diffusive field with five $5\text{km}$ synthetic aperture sources.

The phase of the unsteered diffusive field (Figure

Figure 10. The unwrapped phase of a scalar diffusive field with five lines of synthetic aperture sources. The colorbar displays radians.

Figure 11. The normalized gradient of a scalar diffusive field with five synthetic aperture sources.
Figure 12. The phase-binned gradient of a scalar diffusive field with five unsteered synthetic aperture sources. The movement of the energy is symmetric about both the x and y directions.

The inline and crossline phase-binned gradient plots (Figures 13 and 14) demonstrate how the direction of energy propagation changes with the use of beamforming. The inline field, shown in Figure 13, is steered at a 45° angle in the inline direction which causes constructive interference to occur at large x values compared to the unsteered field in Figure 12, which has the same amount of energy at all values of x. The interference is difficult to view with the log-transform, as previously shown in Figure 4, because the amplitude of the energy is small in the area 45° from the x-axis. The crossline field is steered at 45° which is visible in the lower right panels of Figure 14 as a preferential propagation in the crossline direction (large values of y). Rather than staying symmetric about the y-axis, the crossline steering causes the interference to occur at large y values in late phases. As with the inline field, this is difficult to visualize without the phase-binned display of the direction of propagation. The combination of inline and crossline steering produces a diffusive field with energy in one area of the model. In Figure 15, the concentration of energy to large x and y values is displayed in the later phases of the field. Without the phase-binned gradient technique the change in the direction due to beamform-
Figure 15. The phase-binned gradient of a scalar diffusive field steered in the inline and crossline directions at 45°. The combined inline and crossline steering produces a field whose energy moves out asymmetrically in both the x- and y-directions.

Figure 16. The y-z view of the unsteered diffusive field, panel a), and the combined inline and crossline steered diffusive field, panel b), at a phase of 3π. The oval highlights the change in the direction of the field caused by steering.

The phase-binned gradient of a scalar diffusive field steered in the inline and crossline directions at 45°. The combined inline and crossline steering produces a field whose energy moves out asymmetrically in both the x- and y-directions.

The differences in the direction of the diffusive fields can be used to determine if the synthetic aperture is steered optimally for a specific example. For CSEM fields, where the Poynting vectors are used instead of the gradient the goal is to increase the amount of up-going energy that carries information about the reservoir. This visualization method will identify how the field propagates and how to optimize the beamforming parameters.

4 DISCUSSION AND CONCLUSIONS

The synthetic aperture technique offers a way to address some of the limitations of CSEM without requiring any changes in acquisition. Applying the technique to diffusive fields with 2D source distributions demonstrates the possibilities the technique has to increase the detectability of hydrocarbon reservoirs with CSEM fields. Steering the fields in both the inline and crossline directions causes the strength of the field to move to a localized area. Research is ongoing to apply this technique to synthetic CSEM fields to quantify the benefits of steering with a 2D source. The new visualization technique we introduce demonstrates how a frequency-domain field can appear to propagate as a function of increasing phase. The combined phase and gradient (or Poynting vector) provide a way to visualize how the steering modifies the upgoing field so that amount of information about the target is maximized. This is an improvement over other visualization methods that only show the amplitude or sign of the field. The implementation of these techniques increases the amount of information gleaned from data acquired from the CSEM survey, making CSEM a more valuable tool for industry. The next step is to apply the methods to electromagnetic fields with reservoirs. We will investigate what type of acquisition geometry maximizes the benefits of steering with synthetic aperture and how to optimize the steering.

ACKNOWLEDGMENTS

We are grateful for the financial support from the Shell Gamechanger Program of Shell Research, and thank Shell Research for the permission to publish this work.

References


Laser excitation of elastic waves at a fracture

Thomas Blum(1), Kapser van Wijk(1), Roel Snieder(2), and Mark Willis(3)

(1) Physical Acoustics Laboratory and Department of Geosciences, Boise State University, Boise, ID 83725
(2) Center for Wave Phenomena, Colorado School of Mines, Golden CO 80401, email rsnieder@mines.edu
(3) Formerly at ConocoPhillips Company, Houston, TX 77079, Now at Halliburton, Houston, TX 77032

ABSTRACT
We show that elastic waves can be excited at a fracture inside a transparent sample, by focusing laser light directly onto this fracture. The associated displacement field, measured by a laser interferometer, has pronounced waves that are diffracted at the fracture tips. We confirm that these are tip diffractions from direct excitation of the fracture by comparing them with tip diffractions from scattered elastic waves excited on the exterior of the sample. Being able to investigate fractures – in this case in an optically-transparent material – via direct excitation opens the door to more detailed studies of fracture properties in general.

Key words: scattering, fracture, ultrasound

1 INTRODUCTION

Being able to remotely sense the properties of fractures with elastic waves is of great importance in seismology, e.g. Nakahara et al., (2011) and non-destructive testing, e.g. Larose et al. (2010). For example, in geothermal and hydrocarbon fracturing reservoirs, it is very common to use hydraulic fracturing methods to attempt to increase the native permeability of the rocks above what is present in any naturally occurring fractures. The microseismic events associated with the fracturing process typically radiate seismic energy, which is recorded in nearby wells or at the surface. Much is left to be understood about the nature of such fractures and their relationship to elastic waves, but the scaling issues involved make numerical modeling a challenge. On the other hand, laboratory studies of fractures or faults are used to investigate their mechanical properties, such as stiffness (Pyrak-Nolte and Nolte, 1992), fracture slip-rate, stress drop, or rupture propagation (Nielsen et al., 2008). Typically, fractures under laboratory investigation are either on the surface of samples, or the result of new or growing fractures from an applied stress to induce fracture stick-slip creep (Thompson et al., 2009; Gross et al., 1993). Recently, Blum et al. (2010) used noncontacting techniques to probe a fracture inside a clear sample to recover the fracture compliance. A high-powered laser excites the surface of the sample creating ultrasonic waves. These waves scatter from the fracture and are recorded at the surface of the sample with a laser interferometer (Scruby and Drain, 1990). Here, instead of only exciting the ultrasonic waves at the sample surface, we focus a pulsed infrared (IR) laser beam at the fracture location, turning it into an ultrasonic source. This technique makes it possible to measure the fracture response as a function of source energy, stress on the sample, or the laser beam size and location. By scanning the fracture with a focused IR laser beam it may be possible to measure spatial variations in the fracture properties and delineate barriers and asperities (Scholz, 1990), concepts that are of great importance in earthquake dynamics for example. A localized excitation, along the fracture, could also be used to excite interface waves traveling along the fracture (Roy and Pyrak-Nolte, 1997; Gu et al., 1996) to probe for properties such as fault gouge or the fluids filling the fracture. Here, we illustrate the use of direct excitation of a fracture to investigate the elastic effective size of the fracture by means of tip diffractions. Until now, these are most commonly studied on surface cracks for crack characterization (Masserey and Mazza, 2005).

2 EXPERIMENT

We create a single disk-shaped fracture by focusing a high power Q-switched Nd:YAG laser in a cylinder made of extruded Poly(methyl methacrylate, PMMA), with a diameter of 50.8 mm and a height of 150 mm. The laser generates a short pulse (~ 20 ns) of infrared (IR) light
that is absorbed by the sample material at the focal point and converted into heat. The sudden thermal expansion generates sufficient stress to form a fracture inside the plastic material (Zadler and Scales, 2008; Blum et al., 2011). Anisotropy in the elastic moduli, caused by the extrusion process, results in a fracture with an orientation parallel to the cylindrical axis. The fracture studied here is approximately circular with a diameter of \( \sim 7 \text{ mm} \) (Figure 1).

Elastic waves are excited at the surface of the sample by using the same high-power Q-switched Nd:YAG laser, operated at a much lower power, and with a partially focused beam. When an energy pulse from the laser hits an optically absorbing surface, part of that energy is absorbed and converted into heat. The resulting localized heating causes thermal expansion, which in turn results in elastic waves in the ultrasonic range (Scruby and Drain, 1990).

Typically, the laser is focused on the outside of the sample – but as we explore in this Letter – the laser can also be focused inside the sample. In this case, the planar fracture has a visible contrast with the rest of the sample, seen as a darker region in Figure 1. The Nd:YAG pulsed laser generates energy at a wavelength of 1064 nm, in the near infrared. Therefore, we assume that the optical contrast due to the fracture is also present at the IR wavelength, leading to energy absorption and thermoelastic expansion at the fracture location.

We measure elastic displacement with a laser interferometer, based on a doubled Nd:YAG laser, generating a Constant Wave 250 mW beam at a wavelength of 532 nm. The light is split between a beam reflecting off the sample and one following a reference track inside the sensor. Two-wave mixing of the reflected and reference beams in a photo-refractive crystal delivers a point measurement of the out-of-plane displacement field at the sample surface. The output is calibrated to give the absolute displacement in nanometers (Blum et al., 2010). The frequency response is flat between 20 kHz and 20 MHz, and accurately detects displacements of the order of parts of Angstroms. Since the PMMA sample is transparent for green light, we apply a reflective tape to the surface to reflect light back to the laser receiver.

The location of the non-contacting ultrasonic source and receiver are fixed in the laboratory frame of reference, but the PMMA sample is mounted on a rotational stage. The source-receiver angle \( \delta \) (defined in Figure 2) is therefore constant, here \( \delta = 20^\circ \), and only the orientation of the fracture with respect to the frame of reference, characterized by the angle \( \theta \), changes. Moreover, the source and receiver are focused on the sample in an \( x-y \) plane normal to the cylinder axis (\( z \) axis, Figure 2). While anisotropic, as mentioned above, the extruded PMMA is transversely isotropic, and its elastic properties are therefore invariant with respect to the defined angles of interest.

By computer-controlled rotation of the stage, we measure the elastic field in the \( (x, y) \)-plane for values of \( \theta \) in increments of 1 degree. The signal is digitized with 16-bit precision and a sampling rate of 100 MS/s (mega samples per second) and recorded on a computer acquisition board. For each receiver location, 256 waveforms are acquired and averaged after digitization.

Figure 3 shows the ultrasonic displacement field for the source S1 at the fracture for all recorded azimuths, after applying a 1-5 MHz band-pass filter. As defined in Figure 2, the horizontal axis represents the angle \( \theta \) between the normal to the fracture and the source direction. Electromagnetic interferences are generated by the high-power source laser when the light pulse is emitted, and leads to noise being recorded for short arrival times (0 – 3 \( \mu \)s, highlighted in Figure 2). The arrival
Laser excitation of elastic waves at a fracture

305

θ (°)

Time (µ s)

-90 0 90 180 270

0

10

20

30

40

Source

laser

noise

fP

PfP

fPP

PP

Figure 3. Displacement field generated by excitation of the fracture. fP is the P-wave generated at S1 and traveling directly to the receiver. PfP is the P-wave generated at S2 and scattered by the fracture before reaching the receiver. fPP is the P-wave generated at S1, traveling away from the receiver before bouncing back to the sample surface. Finally, PP is the P-wave generated at S2, traveling across the sample and bouncing back to the receiver.

at approximately 10 µs denoted fP corresponds to the wavefield excited at the fracture. The fPP wave is excited at the fracture and reflects off the backside of the sample.

Next, we apply reflective tape where the source laser beam hits the sample surface at S2, increasing the IR light absorption at the surface and lowering the amount of energy reaching the fracture (Figure 4). We repeat with this configuration the acquisition procedure used in the first experiment (Figure 5). The PfP wave is generated at the surface of the sample, and then scattered by the fracture, while PP is scattered from the backside of the sample. PfP and PP phases are stronger than fP and fPP in Figure 3, confirming that more of the thermoelastic expansion takes place at the surface of the cylinder.

2.1 Fracture tip travel times

The waves fP and PfP in Figures 3 and 5 show a distinct lenticular pattern. For source angles θ = −10° and 170°, the PfP phase is a specular reflection, and the amplitude is a maximum. For intermediate angles, the scattered amplitude decreases (Blum et al., 2011). Note splitting of the wave at intermediate angles into wavelets arriving before and after the specular reflection (see Figure 6). These waves have the travel time and phase of waves diffracted by the crack tips. In particular, for θ = 70°, the receiver is in the plane of the fracture, and therefore the travel time difference between the tips of the fracture the closest and the farthest to the receiver is largest (Figure 4).

Equation 39 in Blum et al. (2011) shows that the P to P scattered amplitude for a planar fracture in a linear-slip model in the Born approximation can be written in the frequency domain as a product of a scaling factor, a factor depending on the mechanical properties of the fracture and the propagation medium, and a form factor that depends on the fracture shape and the wavenumber change from the fracture scattering. Only this last factor carries time information. We show in the Appendix that the corresponding travel times are

\[
t_{\text{tip-sc}} = \frac{R}{a} \left( 2 \pm \frac{a}{R} (\sin \theta (1 + \cos \delta) + \sin \delta \cos \theta) \right),
\]

where a is the radius of the fracture and R the radius of the cylinder. The P-wave velocity is given by
\[ \alpha = 2600 \text{ m/s} \quad \text{(Blum et al., 2011).} \]

Figure 6 shows the \( \text{PfP} \) arrival overlain by the theoretical travel times computed from equation (1) with a fracture radius \( a_{\text{PfP}} = 3.3 \text{ mm.} \)

For the arrival time of the \( \text{fP} \) wave that is excited at the fracture, we consider the geometry of rays originating from the fracture tips and traveling directly to the receiver. The raypaths are shown in Figure 2. Using this geometry the travel time can be expressed as

\[ t_{\text{tip-direct}} = \frac{\sqrt{a^2 \pm 2aR \sin(\theta)} + R^2}{\alpha}. \]  

Due to the fact that the size of the fracture is small compared to the radius of the sample, this travel time is to leading order in \( a/R \) given by

\[ t_{\text{tip-direct}} = \frac{R}{\alpha} \left( 1 \pm \frac{a}{R} \sin(\theta) \right). \]

Figure 7 shows the fracture-source displacement field overlain with the tip arrival time (in blue) computed from equation (3). Just as in Figure 6, the theoretical time for a radius \( a_{\text{fP}} = 3.3 \text{ mm} \) agrees well with the arrival time of the \( \text{fP} \) wave, and the observed size in Figure 1. The good agreement with the visually estimated radius confirms that the whole visually fractured area is mechanically discontinuous and capable of being excited by elastic waves.

3 CONCLUSIONS

Laser-based ultrasonic techniques can not only excite and detect elastic waves at the surface, but can also be used to directly excite heterogeneities (such as fractures) inside an optically transparent sample. This result opens up possibilities for diagnosing the mechanical properties of fractures by directly exciting them. Here, we estimate the effective elastic size of the excited fracture. By scanning the fracture with a focused IR laser beam, it may be possible to measure spatial variations in the fracture properties and delineate barriers and asperities. These concepts are of great importance in earthquake dynamics, although hard to investigate in the field or numerically.

ACKNOWLEDGMENTS

We thank ConocoPhillips for funding this research, Paul Martin for his valuable suggestions, and Randy Nuxoll for his help with sample preparation.

REFERENCES


Nakahara, H., H. Sato, T. Nishimura, and H. Fujiwara, 2011, Direct observation of rupture propagation during the 2011 off the Pacific coast of Tohoku earthquake (m 9.0) using a small seismic array: Earth Planets Space, 63, 589.


the fracture plane, and $J_1$ the first order Bessel function. According to equation (20.53) of Snieder (2009), the asymptotic form of the Bessel function is

$$J_m(x) = \sqrt{\frac{2}{\pi x}} \cos \left( x - (2m + 1)\frac{\pi}{4} \right) + O(x^{-3/2}),$$

(A3)

For the geometry described in Figure 4, the wavenumber change can be expressed as

$$k_1 = \frac{\omega}{\alpha} \left( \sin \theta (1 + \cos \delta) + \sin \delta \cos \theta \right).$$

(A4)

Inserting equations (A3) and (A4) into expression (A2), and expanding the cosine in exponentials gives

$$F(k) \propto e^{i\pi/4} e^{i\omega T} + e^{-i\pi/4} e^{-i\omega T},$$

(A5)

where $T = (a/\alpha) \sin \theta (1 + \cos \delta) + \sin \delta \cos \theta$.

The wavenumber change, $\delta$, quantifies the delay time of the tip diffraction arrivals relative to the arrival time $t = 2R/\alpha$ for a ray reflecting at the center of the fracture. Therefore, the total tip diffraction travel times for the scattered arrival are given by equation (1). Note that this expression predicts a phase shift $\exp(\pm i\pi/4)$ for these waves that is characteristic of edge-diffracted waves (Keller, 1978).

**APPENDIX A: TIP-DIFFRACTION TIMES FROM FORM FACTOR**

Equation 39 of Blum et al. (2011) shows that the P to P scattered amplitude for a planar fracture in a linear-slip model in the Born approximation is

$$f_{P,P}(\hat{n};\hat{m}) = \frac{\omega^2}{4\pi\rho a^4} AF(k_{\alpha}(\hat{n} - \hat{m}))$$

$$\times \left\{ \lambda^2 \eta_N + 2\mu \eta_N \left( (\hat{n} \cdot \hat{f})^2 + (\hat{m} \cdot \hat{f})^2 \right) + 4\mu^2(\eta_N - \eta_T)(\hat{n} \cdot \hat{f})(\hat{m} \cdot \hat{f}) \right\} \left( A_1 \right)$$

where $\omega$ is the angular frequency, $\alpha$ the P-wave velocity, $\rho$ the density of the material, $\lambda$ and $\mu$ the Lamé parameters, $A$ the surface area of the fracture, and $\eta_N$ and $\eta_T$ the normal and tangential compliances, respectively, for the linear-slip model. The unit vectors $\hat{n}$ and $\hat{m}$ denote the directions of incoming and outgoing waves, respectively, and $\hat{f}$ is the unit vector normal to the fracture (see Figure 4).

The prefactor $(\omega^2/4\pi\rho a^4)A$ does not carry time information. The factor in curly brackets contains the angular dependence of the scattering amplitude, and depends only on the mechanical properties of the fracture $\eta_N$ and $\eta_T$ of the sample material, and on the directions of the incoming and outgoing waves relative to the fracture orientation. The form factor $F(k_{\alpha}(\hat{n} - \hat{m}))$ depends on the fracture size and shape, and contains travel time information. For the case of a circular fracture, the form factor is given by eq. (33) of Blum et al. (2011):

$$F(k_{\alpha}(\hat{n} - \hat{m})) = \frac{2}{k_{||} a} J_1(k_{||} a),$$

(A2)

where $a$ is the radius of the fracture, $k_{||}$ the projection of the wavenumber change during the scattering onto the fracture plane.
Analyzing the coda from correlating scattered surface waves

T. Dylan Mikesell(1), Kasper van Wijk(1), Thomas E. Blum(1), Roel Snieder(2), and Haruo Sato(3)

(1) Physical Acoustics Laboratory and Department of Geosciences, Boise State University, Boise, ID 83725
(2) Center for Wave Phenomena, Colorado School of Mines, Golden CO 80401, email rsnieder@mines.edu
(3) Department of Geophysics, Graduate School of Science, Tohoku University, Aoba-ku, Sendai, Miyagi, 980-8578, Japan

ABSTRACT
The accuracy of scattered Rayleigh waves estimated using an interferometric method is investigated. Summing the cross correlations of the wave fields measured all around the scatterers yields the Green’s function between two excitation points. This accounts for the direct wave and the scattered field (coda). The correlations themselves provide insights into the location of the scatterers as well as which scatterer is responsible for particular parts of the coda. Furthermore, these measurements confirm a constant-time arrival in the correlations, not part of the Green’s function, but which has previously been derived as a result of the generalized optical theorem.

Key words: seismic interferometry, ultrasound, coda, spurious arrivals

1 INTRODUCTION
Numerical and laboratory experiments have shown how cross correlations of field fluctuations recorded at two points in open and closed systems provide an estimate of the anti-causal and causal Green’s functions between these points (Lobkis and Weaver, 2001; Roux and Fink, 2003; Malcolm et al., 2004; Derode et al., 2003; van Wijk, 2006). Wapenaar and Fokkema (2006) showed how a reciprocity theorem relaxes the necessity for sources of the field fluctuations to be present throughout the volume to sources on a surface surrounding these points. These insights have made a large impact on countless ocean acoustic ambient noise studies e.g., Roux et al. (2004); Traer et al. (2011); Siderius (2012), but actually dates back to Aki (1957). Dispersive ballistic surface waves extracted by cross correlation of ambient noise measurements are now a major tool for imaging the Earth’s crust (Campillo and Paul, 2003; Shapiro et al., 2005; Sabra et al., 2005). Moreover, coda waves derived from these correlations have been used for subsurface monitoring (Sens-Schönfelder and Wegler, 2006; Brenguier et al., 2008). We describe here how the coda can be exploited in this correlation technique to locate individual scatterers, and we investigate the accuracy of the recovered coda using an approximate acoustic method rather than the complete elstodynamic Green’s function retrieval method (Wapenaar, 2004).

We estimate the ultrasonic Rayleigh-wave Green’s function for surface waves propagating in an aluminum block with 15 scatterers at the surface. In the far field, the acoustic Green’s function $G$ for a medium with constant velocity $c$ and density $\rho$ can be approximated by summing cross correlations of the measured pressure field $u$ (equation 32 of Wapenaar and Fokkema (2006)):

$$ 2i\omega \rho c \oint_{\partial D} u^*(x_A, x, \omega) u(x_B, x, \omega) dl, $$

where $G(x_A, x_B, \omega)$ is the vertical component of the Rayleigh-wave Green’s function at $x_A$ from a source at $x_B$, and $Im$ denotes the imaginary part. $S(\omega)$ is the power spectrum of the sources on the closed contour $\partial D$. Strictly speaking, expression (1) is only valid for acoustic media; however, Halliday and Curtis (2008) show that $G$ is equivalent to the $G_{zz}$ component of the elastic surface wave Green’s tensor when $c$ is the Rayleigh-wave speed and $u$ the vertical component of the displacement. We refer to expression (1) as the acoustic approximation for Green’s function retrieval. Transforming expression (1) to the time domain states that the sum of the cross correlations of the wave fields from sources on...
the boundary $\partial D$ is proportional to the causal and anti-causal Green’s function that describes wave propagation between $x_A$ and $x_B$.

2 LABORATORY MODEL AND ULTRASONIC DATA

One side of an aluminum block (280 mm $\times$ 230 mm $\times$ 215 mm) contains cylindrical holes 1 mm in diameter and 3 cm deep (Fig. 1(a)). A high-powered pulsed Nd:YAG laser generates ultrasonic waves by briefly (15 ns) heating a 1 mm wide circular spot on the surface. The heating causes thermoelastic expansion and generates broadband ultrasonic waves having a central frequency of 600 kHz. The Rayleigh-wave velocity in the aluminum is $c = 2.9$ mm/$\mu$s, corresponding to a dominant wavelength of $\sim$ 5 mm. We measure the vertical component of the displacement as a function of time at a point on the surface of the cube with a laser interferometer, based on a constant-wave 250 mW Nd:YAG laser at 532 nm. The receiver uses two-wave mixing in a photorefractive crystal and is calibrated to measure the absolute out-of-plane displacement field. The measured response is flat between 20 kHz and 20 MHz, with sensitivity on the order of parts of angstroms (see Blum et al. (2010) for a complete description).

With the source laser at $x_A$ (star), we record the wave field at each receiver position (triangles) for 60 $\mu$s at a sample rate of 20 MS/s. There are 412 receivers in total, spaced $\sim$ 1 mm apart surrounding the source and scatterers. At each receiver position, we repeat the source excitation 128 times and average the recordings to increase the signal-to-noise ratio. The displacement wave fields are presented in Fig. 1(b), after we apply a cosine taper to the first 5 $\mu$s to suppress electronic noise from the laser recording system and a band-pass filter between 100-2000 kHz. We mute reflections from the edge of the block, when these arrive before $t = 60 \mu$s.

The wave fields are displayed as a function of receiver azimuth ($0 \leq \theta < 360^\circ$), defined with respect to the center of the receiver array, with positive $\theta$ in the clockwise direction and $\theta = 0$ in the upper left corner of the square. The arrival with the largest amplitude is the direct Rayleigh wave, followed by scattered Rayleigh waves with the opposite polarity caused by scattering at the air-filled holes. The sharp variations in the coherent arrivals are due to the rectangular nature of the underlying acquisition geometry.

2.1 Auto correlation

According to expression (1), the Green’s function between two sources can be estimated from receivers on $\partial D$ based on reciprocity of the wave equation (van Maren et al., 2005; Curtis et al., 2009). Here we estimate the vertical component of the full displacement wave field – both direct and scattered waves – between two sources, because with our acquisition setup it is easier to record with many receivers than to use many sources. The auto correlation of the wave field for each receiver as excited by the source at $x_A$ is shown in Fig. 2(a), which henceforth we call a correlation gather. The sum of these auto correlations is shown in Fig. 2(b). As we have not taken the time derivative as defined in expres-
Figure 2. (a) The auto correlation gather of all receivers for a source at $x_A$. The stars indicate the stationary-phase points for scatterer 1 and the dashed lines trace out the correlation of the scattered wave and direct wave along the full receiver array. (b) The sum of the correlations, amplified 20 times the maximum amplitude. (c) The stationary-phase receiver ($\theta = 34^\circ$) aligns with $x_A$ and scatterer 1. The arrival time $t \approx -26 \mu s$ corresponds to a scatterer at a distance $r = 37.5$ mm from $x_A$.

2.2 Cross correlation

To determine the accuracy of the coda retrieval, and because it is not trivial to measure the impulse response directly for a coinciding source and receiver position, we repeat the same acquisition described above for a source at $x_B$ (Fig. 1(a)). In this case, between $\theta = 60^\circ$ and $180^\circ$ the side reflection from the block edge interferes with waves scattered from some of the cylindrical holes. Muting the side reflection also removes some of the scattering coda. Next we cross correlate the processed wave fields recorded at each receiver from the two sources (Fig. 3(a)) and the sum over all sources (Fig. 3(b)). Following expression (1), each trace in the sum is weighted by the difference in angle between the receiver and proceeding receiver.

Toward estimating the wave field that propagates between $x_A$ and $x_B$, we time-reverse the anticausal weighted sum and stack with the causal weighted sum to improve the signal-to-noise ratio. As prescribed by expression (1), we then take a time derivative. Because the correlation contains a contribution $S$ from the source wavelet, we deconvolve a tapered version of the direct Rayleigh-wave arrival from our result using a water-level deconvolution Snieder and Şafak (2006). The tapered wavelet used in the deconvolution is shown in the inset of Fig. 3(c).

We compare our estimate of the wave field propagating between a source $x_A$ and a virtual receiver $x_B$ to a direct measurement for a source at $x_A$ and a real receiver at $x_B$ (Fig. 3(c)). The amplitudes of the two wave fields are normalized with the maximum of the direct Rayleigh wave. The direct surface wave is extracted with great accuracy. The amplitude of the coda — relative to the direct wave — is accurately estimated, and some arrivals in the coda are well reproduced, while others are not. Differences between the direct estimate of the impulse response and the result based on expression (1) are caused by violations of the assumptions that went into the approximate equation (1). An exact formulation for the extraction of the full elastodynamic Green’s function has been developed (Wapenaar, 2004). This formulation requires, however, multicomponent point forces and double couples that excite the field fluctuations, and is difficult to implement in laboratory...
and field experiments. Specifically in this experiment, we 1) suppress some scattered waves when muting the side reflection, 2) violate the far-field assumption, 3) apply an approximate deconvolution, and 4) measure the wave field at a finite number of locations on $\partial D$ rather than everywhere. Halliday and Curtis (2008) justified the use of this acoustic formulation based on its accuracy for extracting the direct surface wave, not necessarily on scattered arrivals. One aspect of recovery of the coda in this technique is that we are able to identify which scatterer is responsible for a certain time-section of the coda.

3 WAVES WITH A CONSTANT ARRIVAL TIME IN THE CORRELATION GATHER

In theory – and previously only shown numerically – the accurate extraction of scattered waves from the correlation of field fluctuations relies on the cancellation of correlated energy with a constant arrival time. These waves come from the correlation of waves traveling from a source to a particular scatterer and on to both receivers; the arrival time of this phase is independent of source location (Snieder et al., 2008). Its cancellation is due to the generalized optical theorem that constrains the scattering coefficient. Figure 3(a) does not provide any evidence of correlations with a constant arrival time because of the large amplitude of the direct wave. In Fig. 4(a), the direct waves are muted before cross correlation. In other words, it is the correlation gather for scattered waves only. Several events with a constant arrival time are now visible, some of these are highlighted with boxes around $t \sim -9 \mu s$ and $+5 \mu s$. The scattering coefficient of surface waves depends, in general, strongly on the scattering angle (Snieder, 1986), and the events with a constant arrival time in Fig. 4(a) are strongest when the source, receiver and scatterer are aligned. Because there are 15 scatterers in the model, there should be 15 waves with a constant arrival time. It is difficult to verify this because the correlation gather is complicated by cross terms of waves radiated by different scatterers (whose arrival time depends on the receiver angle $\theta$).

The wave field from the summed correlation gather of only the scattered waves is shown in Fig. 4(b). Note that in contrast from the wave field extracted from the cross correlation of the full wave field (Fig. 3(b)), Fig. 4(b) does not show the direct wave at all. In fact, the summed wave field is dominated by arrivals $t \sim -9 \mu s$ and $t \sim +5 \mu s$ that correspond to the sum of the waves in the boxes in the correlation gather with a constant arrival time. These spurious arrivals should not contribute to the estimated impulse response, and theory and numerical examples show that these waves cancel by stationary phase contribution from a cross term of direct and scattered waves (Snieder et al., 2008). Since we mute the direct wave, such cancellation does not take place, with the result that these spurious arrivals dominate in Fig. 4(b). It has been shown theoretically that in order to retrieve the scattered waves, one needs the cross correlation of the direct wave with scattered waves (Snieder and Fleury, 2010). Since the direct waves are muted in the correlation gather of Fig. 4(a), one cannot expect that the waves in Fig. 4(b) accurately represent the scattered waves.

4. CONCLUSION

We present laboratory measurements of scattered ultrasonic surface waves that propagate in an aluminum
Figure 4. (a) Coherent events (highlighted in boxes) appear in the cross-correlation gather at constant times $t \sim -9 \mu s$ and $t \sim +5 \mu s$ after suppressing the direct-wave energy. These events are strongest in the forward scattering direction and contribute the most energy to the sum (b).

block with cylindrical holes that act as scatterers. An estimate of the wave field between two source locations can be obtained by summing the cross correlated wave fields measured at receiver locations enclosing the sources and scatterers. The correlation gather computed from the recorded wave fields shows distinct arrivals related to individual scatterers, and we show that the stationary-phase points of these arrivals can be used to constrain the location of scatterers.

The accuracy of our Green’s function estimate can be compared by directly measuring the response at the location of one of the two sources. The direct wave is reconstructed quite accurately, but the reconstruction of the coda is less accurate. The acoustic formulation for Green’s function retrieval we use is justified on arguments that apply to the direct wave only, and we attribute the reduced ability to extract the coda to the inaccuracy of using the acoustic formulation for scattered elastic surface waves.

Finally, theory predicts that for accurate recovery of the wave field, the correlation gather should contain energy with a constant arrival time from the correlation of scattered arrivals from the same scatterer at each receiver. We show the presence of such arrivals in laboratory data by muting the direct wave before cross correlation of the wave fields. Summing this cross-correlation gather does not lead to a reconstruction of either the direct or scattered waves. This agrees with theory, which predicts that the cross correlation of direct and scattered waves are essential in Green’s function extraction by cross correlation.

REFERENCES


Snieder, R., 1986, 3-D linearized scattering of surface
Estimating near-surface shear-wave velocities in Japan by applying seismic interferometry to KiK-net data

Nori Nakata & Roel Snieder
Center for Wave Phenomena, Geophysics Department, Colorado School of Mines, Golden, Colorado 80401

ABSTRACT
We estimate shear-wave velocities in the shallow subsurface throughout Japan by applying seismic interferometry to the data recorded with KiK-net, a strong-motion network in Japan. Each KiK-net station has two receivers; one receiver on the surface and the other in a borehole. By using seismic interferometry, we extract the shear wave that propagates between these two receivers. Applying this method to earthquake-recorded data at all KiK-net stations from 2000 to 2010 and measuring the arrival time of these shear waves, we analyze monthly and annual averages of the near-surface shear-wave velocity all over Japan. Shear-wave velocities estimated by seismic interferometry agree well with the velocities obtained from logging data. The estimated shear-wave velocities of each year are stable. For the Niigata region, we observe a velocity reduction caused by major earthquakes. For stations on soft rock, the measured shear-wave velocity varies with the seasons, and we show negative correlation between the shear-wave velocities and precipitation. We also analyze shear-wave splitting by rotating the horizontal components of the surface sensors and borehole sensors and measuring the dependence on the shear-wave polarization. This allows us to estimate the polarization with the fast shear-wave velocity throughout Japan. For the data recorded at the stations built on hard-rock sites, the fast shear-wave polarization directions correlate with the direction of the plate motion.

Key words: seismic interferometry, shear-wave splitting, shear-wave velocity, precipitation, near surface, KiK-net

1 INTRODUCTION
Seismic interferometry is a powerful tool to obtain the Green's function that describes wave propagation between two receivers (e.g., Aki, 1957; Claerbout, 1968; Lobkis and Weaver, 2001; Roux and Fink, 2003; Schuster et al., 2004; Wapenaar et al., 2004; Bakulin and Calvert, 2006; Snieder et al., 2006; Wapenaar and Fokkema, 2006). Seismic interferometry is applied to ambient noise (e.g., Hohl and Mateeva, 2006; Draganov et al., 2007, 2009; Brenguier et al., 2008; Strehly et al., 2008; Lin et al., 2009), traffic noise (e.g., Nakata et al., 2011), production noise (e.g., Miyazawa et al., 2008; Vasconcelos and Snieder, 2008), earthquake data (e.g., Larose et al., 2006, Sens-Schönfelder and Wegler, 2006; Snieder and Şafak, 2006; Ma et al., 2008; Ruigrok et al., 2010), and active sources (e.g., Bakulin and Calvert, 2004; Mehta et al., 2008).

In Japan, large seismometer networks, such as Hi-net, F-net, K-NET, and KiK-net (Okada et al., 2004), are deployed. By using these networks for seismic interferometry, Tonegawa et al. (2009) extract the deep subsurface structure of the Philippine Sea slab. These data have also been used to observe time-lapse changes in small regions (Wegler and Sens-Schönfelder, 2007; Sawazaki et al., 2009; Wegler et al., 2009; Yamada et al., 2010). Each KiK-net station has two receivers, one on the ground surface and the other at the bottom of a borehole. One can estimate the body-wave velocity between two receivers by using seismic interferometry (Trampert et al., 1993; Snieder and Şafak, 2006; Mehta et al., 2007b; Miyazawa et al., 2008).

By applying seismic interferometry to KiK-net data, we analyze near-surface velocities throughout Japan. Because KiK-net has recorded strong-motion seismograms continuously since the end of 1990s, the data are available for time-lapse measurements. Measuring time-lapse changes of the shallow subsurface is important for civil engineering and for estimating the site response to earthquakes. Previous studies ex-
tracted time-lapse changes caused by earthquakes (Li et al., 1998; Vidale and Li, 2003; Schaff and Beroza, 2004; Wegler and Sens-Schönfelder, 2007; Brenguier et al., 2008). Interferometry applied to a single KiK-net station has also been used to measure time-lapse change due to earthquakes (Sawazaki et al., 2009; Yamada et al., 2010). Interferometric studies have shown changes in the shear-wave velocity caused by precipitation (Sens-Schönfelder and Wegler, 2006) and have measured shear-wave splitting (Bakulin and Mateeva, 2008; Miyazawa et al., 2008). We study the annual and monthly averages of the shear-wave velocity and the fast shear-wave polarization directions for stations all over Japan, and the temporal change in shear-wave velocity in the Niigata prefecture for three major earthquakes.

This paper presents data processing of KiK-net data with seismic interferometry. We first introduce the properties of KiK-net. Next, we show the data analysis method. Then we present near-surface shear-wave velocities in every part of Japan. Finally, we interpret these velocities to study time-lapse changes, which are related to major earthquakes and precipitation, and present measurements of shear-wave splitting.

2 KIK-NET DATA

About 700 KiK-net stations are distributed in Japan (Figure 1). The stations are operated by the National Research Institute for Earth Science and Disaster Prevention (NIED). Each station has a borehole and two seismographs which record strong motion at the bottom and top of the borehole. Each seismograph has three components: one vertical component and two horizontal components. Although the two horizontal components of the surface seismograph are oriented in the north-south and east-west directions, respectively, the horizontal components of the borehole seismograph are not always aligned with the north-south and east-west directions because of technical limitations. Therefore, we rotate the directions of the borehole seismograph north-south and east-west directions before data processing. The depth of about 25% of the boreholes is 100 m, and the other boreholes are at greater depth. Since our target is the near surface, we use the stations with a depth less than 525 m, which accounts for 94% of the stations. The sampling interval is either 0.005 or 0.01 s, depending on the station and the recording date.

We show example records of an earthquake in Figure 2. Figure 2a illustrates bandpass-filtered time series, and Figure 2b the power spectra of the unfiltered data. As shown in Figure 2b, most energy is confined to 1-13 Hz, and we apply a bandpass filter over this frequency range for all data processing. In Figure 2, UD denotes the vertical component, NS the north-south direction horizontal component, and EW the east-west direction horizontal component. In Figure 2a, the P wave arrives
shear-wave velocity between these sensors as determined in this study. A bias in the velocity estimation due to non-vertical propagation depends on the deviation of \( \cos \theta \) from its value for vertical incidence, \( \cos 0^\circ = 1 \).

### 3 RETRIEVAL OF THE WAVEFIELD BETWEEN RECEIVERS

We apply seismic interferometry to the recorded earthquake data of each station for retrieving the wavefield where the borehole receiver behaves as a virtual source. Several algorithms have been used in seismic interferometry to obtain the wavefield. These include cross correlation (e.g., Claerbout, 1968; Bakulin and Calvert, 2004), deconvolution (e.g., Trampert et al., 1993; Snieder and Şafak, 2006), cross coherence (e.g., Aki, 1957; Prieto et al., 2009), and multidimensional deconvolution (e.g., Wapenaar et al., 2008; Minato et al., 2011).

We introduce the cross-correlation and deconvolution algorithms. We denote the wavefield, excited at source location \( s \) that strikes the borehole receiver at location \( b \) by \( u(r_b, s, \omega) = S(r_b, s, \omega) \), where \( S(r_b, s, \omega) \) is the incoming wavefield that includes the source signature of the earthquake and the effect of propagation such as attenuation and scattering, in the frequency domain. The corresponding wavefield recorded at the surface receiver at location \( r_s \) is given by

\[
    u(r_s, s, \omega) = 2G(r_s, r_b, \omega)S(r_b, s, \omega),
\]

where the factor 2 is due to the presence of the free surface at \( r_s \). Because the wavefield striking the borehole receiver is close to a vertically propagating plane wave, \( G(r_s, r_b, \omega) \) is the plane wave Green’s function that accounts for the propagation from the borehole seismometer to the surface seismometer.

The cross-correlation approach to retrieve the wavefield in one dimension is given by Wapenaar et al. (2010)

\[
|S(r_b, s, \omega)|^2 G(r_s, r_b, \omega) = \frac{2j\omega}{pc} u(r_s, s, \omega) u^*(r_b, s, \omega),
\]

where \( \rho \) is the mass density of the medium, \( c \) the wave propagation velocity, \( j \) the imaginary unit, and \( * \) the complex conjugate. The regularized deconvolution, which is similar to cross correlation, is given by

\[
G(r_s, r_b, \omega) = \frac{u(r_s, s, \omega)}{u(r_b, s, \omega)} \approx \frac{u(r_s, s, \omega) u^*(r_b, s, \omega)}{|u(r_b, s, \omega)|^2 + \epsilon},
\]

where \( \epsilon \) is a regularization parameter (Mehta et al., 2007b,a). The deconvolution is potentially unstable due to the spectral division, and we avoid divergence by adding a positive constant \( \epsilon \) to the denominator (equation 3). Note that the deconvolution eliminates the
imprint of waveform $S(r_b, s, \omega)$, which is incident on the borehole receiver. We derive the features of cross-correlation and deconvolution interferometry in Appendix A.

4 DATA PROCESSING

We use 111,934 earthquake-station pairs that are recorded between 2000 and 2010. The magnitude range is confined between 1.9 and 8.2. The cosine of the angle of incidence $\cos \theta$ of the wave propagating between the receivers at each station is greater than 0.975, even for the events that are the furthest away. The bias introduced by non-vertical propagation thus is less than 2.5%, and for most measurements it is much smaller. First, we check the data quality and drop some seismograms by a visual inspection using the signal-to-noise ratio as a criterion. Additionally, we discard stations with a borehole seismometer at a depth greater than 525 m because we focus this study on the near surface. We remove the DC component of the data by subtracting the average of each seismogram. After stacking, we estimate the arrival time by seeking the three adjacent samples with the largest values and apply quadratic interpolation to find the time at which the deconvolved data have a maximum amplitude (Figure 3). This time is the travel time for a shear wave that propagates between the borehole and surface sensors. We use this travel time to compute the shear-wave velocity of the region between the two receivers.

4.1 Estimating the shear-wave velocity

Before we apply annual stacking or monthly stacking, we resample the data from 0.005-s interval to 0.01-s interval if the data that are stacked include both 0.005-s and 0.01-s sampling-interval data. After stacking, we estimate the arrival time by seeking the three adjacent samples with the largest values and apply quadratic interpolation to find the time at which the deconvolved data have a maximum amplitude (Figure 3). This time is the travel time for a shear wave that propagates between the borehole and surface sensors. We use this travel time to compute the shear-wave velocity of the region between the two receivers.

4.2 Computing the average and standard deviation of the velocity of the annual or monthly stacks

To interpret time-lapse variations in the velocities, we need to compute the average and standard deviation of the velocities within a region. Let us denote the estimated velocity by $v_i(m, y)$, where $v_i$ is the shear-wave velocity at station $i$, in month $m$, and year $y$. This velocity is already averaged over each month. Each station has a different velocity. In order to quantify the time-lapse variations of the velocity, we subtract the average value of each station before calculating temporal variation in the annual or monthly average:

$$\Delta v_i(m, y) = v_i(m, y) - \bar{v}_i,$$ (4)

where $\bar{v}_i$ is an average velocity of station $i$ over all months and years. Then we compute either the annual or monthly average of the velocity variation over stations $\Delta v$, and we also compute the standard deviation of this quantity.
Near-surface S-wave velocities in Japan

Figure 4. Annual-stacked wavefields by using (a) cross correlation and (b) deconvolution interferometry at station NIGH13. The surface and borehole receiver orientation directions are north-south. Epicenter locations are illustrated in Figure 5.

Figure 5. Epicenters used in Figure 4 from 2000 to 2004 (a) and from 2005 to 2010 (b) at station NIGH13. At the right in each panel, we show the number of earthquakes we use to obtain the waveforms in Figure 4 in each year. The size of each circle refers the magnitude of each earthquake and the color indicates the depth. The white triangle illustrates the location of station NIGH13. Because of the proximity of events, many circles overlap.

4.3 Comparison between cross-correlation and deconvolution interferometry

We compare the cross-correlation and deconvolution approaches using the annual-stacked wavefields (Figure 4). We show the locations of the epicenters of the used earthquakes in Figure 5. The annual stacks of the waveforms obtained by cross correlation are shown in Figure 4a. These waveforms are not repeatable from year to year and often do not show a pronounced peak at the arrival time of the shear wave at around $t = 0.15$ s. We attribute the variability in these waveforms to variations in the power spectrum $|S(r_b, s, \omega)|^2$ of the waves incident at the borehole receiver (equation A1). In contrast, the annual stacks of the waveforms obtained by deconvolution shown in Figure 4b are highly repeatable and show a consistent peak at the arrival time of the shear wave. The consistency of these waveforms is due to the deconvolution that eliminates the imprint of the incident wave $S(r_b, s, \omega)$ (equation A2). Consistent with
earlier studies (Trampert et al., 1993; Snieder and Şafak, 2006), we use deconvolution to extract the waves that propagate between the seismometers at each KiK-net station.

4.4 Shear-wave splitting

We investigate shear-wave splitting by measuring the shear-wave velocity as a function of the polarization. We rotate both surface and borehole receivers from 0 to 350 degrees using a 10-degree interval. The north-south direction is denoted by 0 degrees, and the east-west direction by 90 degrees. Because a rotation over 180 degrees does not change the polarization, the 0- to 170-degree wavefields are the same as the 180- to 350-degree data. We apply deconvolution interferometry to the rotated wavefields, located at the surface and borehole receivers with the same polarizations, for determining the velocity of each polarization. Because the velocity for each polarization is related to the velocities of the fastest and slowest shear waves (Appendix B), we can estimate shear-wave splitting from the velocity difference. We cross-correlate the deconvolved wavefield for every used polarization (from 0 to 350 degrees in 10-degree intervals) with the deconvolved wavefield obtained from the motion in the north-south direction. This allows us to quantify the polarization dependence of the shear-wave velocity. Similar to the process described in section 4.1, we compute annual stacks of cross-correlated wavefields and pick the peak amplitudes of stacked wavefields by using quadratic interpolation.

We can separate the velocity $v(\phi)$ as a function of polarization direction $\phi$ into the isotropic and anisotropic terms using a Fourier series expansion (Appendix C):

$$v(\phi) = v_0 + v_1 \cos 2\phi + v_2 \sin 2\phi.$$  (5)

In this expression, $v_0$ is the isotropic component of the velocity, and $\sqrt{v_1^2 + v_2^2}$ the anisotropy. We assume the splitting time to be much smaller than the period of the wavefield. Because the wavefields of each polarization data are symmetric by 180 degrees, the anisotropy depends on polarization through a dependence of $2\phi$.

5 RETRIEVED NEAR-SURFACE SHEAR-WAVE VELOCITIES IN JAPAN

Using deconvolution interferometry at each station, we obtain the wavefield that corresponds to a plane wave propagating in the near-vertical direction ($\cos \theta > 0.975$) between the borehole receiver and surface receiver at each station. In this section, we show the wavefields of the annual stack, monthly stack, and shear-wave splitting.

Figure 6. Annual-stacked wavefields (curves) with the interpolated largest amplitude (circles) at station NIGH13. The horizontal line at around 0.15 s is the shear-wave arrival time determined from logging data. From left to right, we show annual stacks from 2000 to 2010. The source and receiver polarization directions are the north-south direction.

5.1 Annual and monthly stacks

Figure 6 shows the annual-stacked wavefields at station NIGH13 represented by the large black circle in Figure 1. At this station, the sampling interval is 0.005 s until 2007 and is 0.01 s after 2008. In Figure 6, the deconvolved wavefields have good repeatability and a pronounced peak amplitude. After we apply quadratic interpolation (section 4.1), the determined arrival times (the black circles in Figure 6) correlate well with the travel time which is obtained from logging data (the horizontal line in Figure 6). The logging data is measured using a logging tool and by Vertical Seismic Profiling (VSP). The seismic source of VSP is a vertical-component vibrator. For finite offsets this source generates shear waves. We determine the average velocity from the logging data by computing the depth average of the slowness, because this quantity accounts for the vertical travel time. Because of the quadratic interpolation, the measured travel times show variations smaller than the original sampling time.

After determining the arrival times of all stations, we compute the shear-wave velocities by using the known depth of the boreholes. Applying triangle-based cubic interpolation (Lawson, 1984) between stations, we create the shear-wave velocity map of Japan in each year (Figure 7). To reduce the uncertainty of the velocity estimation we use only the stations which give deconvolved waves with an arrival time greater than 0.1 s.
Thus, we obtain the near-surface shear-wave velocities throughout Japan by applying seismic interferometry to KiK-net data. The shear-wave velocity obtained from logging data is shown in the top left in Figure 7. Note that the velocities measured in different years are similar. In Figure 8, we crossplot the velocities estimated by interferometry in 2008 and obtained from logging data. The data are concentrated along the black line, which indicates the degree of correlation between the shearwave velocity obtained from logging data and from seismic interferometry.

We also analyze seasonal changes and show the monthly-stacked wavefields at station NIGH13 in Figure 9. The monthly stacked wavefields also have good repeatability between different months.

5.2 Shear-wave splitting

In Figure 10a, we show the wavefields of the shear-wave splitting analysis at station NIGH13 in 2010 that are obtained by the sequence of deconvolution and cross correlation described in section 4.4. Each trace is plotted at the angle that is equal to the shear-wave polarization used to compute that trace. The thick solid line in Figure 10a shows the interpolated maximum amplitude time of each waveform. The dashed circle shows the arrival time for the wave polarized in the north-south direction. For the polarizations where the thick solid line is outside of the dashed circle, the shear-wave velocity is slower. The fast and slow shear polarization directions in Figure 10a are 22 degrees and 71 degrees clockwise from the north-south direction, respectively. The angle between these fast and slow directions is 93 degrees, which is close to 90 degrees as predicted by theory (Crampin (1985) and Appendix C). The 3-degree discrepancy could be caused by data noise or discretization errors. At station NIGH13 in 2010, the fast polarization shear-wave velocity \( v_{\text{fast}} \) is 638 m/s, the slow velocity \( v_{\text{slow}} \) is 593 m/s, and the anisotropy parameter \( (v_{\text{fast}} - v_{\text{slow}})/v_{\text{fast}} \) is 7% (see Figure 10b). The dif-
6 INTERPRETATION OF SHEAR-WAVE VELOCITIES AND SHEAR-WAVE SPLITTING

6.1 Influence of major earthquakes

The near-surface shear-wave velocity in Japan is similar between years (see Figure 7), which means the near-surface structure is basically stable. In this section, we focus on a small region. We use Δv (calculated by the method presented in section 4.2) and the fast shear-wave polarizations shown in Figure 11 to analyze the influence of major earthquakes in the Niigata prefecture (the dark shaded area in Figure 1). Three significant earthquakes, shown by the dashed arrows in each panel, occurred during the 11 years.

Figure 11a shows the velocity variation Δv for the isotropic component v0 computed by equation 5 compiled over periods one year before and three months after the major earthquakes. We use all stations in the Niigata prefecture and compute the average over the stations. Each box depicts the time range (horizontal extent) and the error in the average velocity over that time interval (vertical length). The error in the velocity is given by the standard deviation of measurements from different earthquakes in each time interval (section 4.2). In Figure 11a, these average velocities show significant velocity reduction after the major earthquakes. The average isotropic velocity of all stations in the region from 2000 to 2010 is 662 m/s, and the relative velocity change of each earthquake is around 3-4%. Similar velocity variations caused by major earthquakes were reported earlier; for example, Sawazaki et al. (2009) analyze the variations caused by the 2000 Western-Tottori Earthquake, Yamada et al. (2010) analyze the variations caused by the 2008 Iwate-Miyagi Nairiku earthquake using KiK-net stations, while Nakata and Snieder (2011) observe a velocity reduction of about 5% after the 2011 Tohoku-Oki earthquake. To increase the temporal resolution of the velocity change, we compute velocity changes averaged over periods one year before and three months after the major earthquakes (Figure 11a) because Sawazaki et al. (2009) found that the velocity reduction is sustained over a period of at least three months after an earthquake.

The stations on soft-rock sites have a greater velocity reduction than those on hard-rock sites. (We define soft- and hard-rock sites from the estimated shear-wave velocity; hard-rock sites have a shear-wave velocity greater than 600 m/s, while soft-rock sites have a shear-wave velocity less than 600 m/s.) For the used event-station pairs, the velocity reduction does not change measurably with the distance from the epicenter. This is an indication that the velocity reduction depends
Near-surface S-wave velocities in Japan

Figure 10. (a) Cross correlograms along every 10-degree polarization direction in 2010 at station NIGH13. Each trace is plotted at an angle equal to the polarization direction used to construct that trace. The dashed circle indicates the peak-amplitude time of the north-south direction, and the thick solid line represents the peak-amplitude time for each polarization direction. (b) Shear-wave velocities computed from the thick solid line in panel (a). Black circles represent the quadratic-interpolated fast and slow polarization shear-wave velocities.

Figure 11. (a) The isotropic component of the shear-wave velocity (Fourier coefficient $v_0$) averaged over one year before and three months after three major earthquakes. (b) The direction of the fast shear-wave polarization averaged over same intervals. All data are computed for stations in the Niigata prefecture (the dark shaded area in Figure 1). In each panel, the label of the year is placed in the middle of each year. The dashed arrows in panel (a) and the vertical dashed lines in panel (b) indicate the times of the three major earthquakes shown with the black crosses in Figure 1. The numbers at the left and right of dashed arrows (a) and lines (b) are the number of earthquakes we use for determining velocities and polarizations before and after the major earthquakes, respectively. In panel (a), each velocity is the velocity variation $\Delta v$ in equation 4. The horizontal extent of each box depicts the time interval used for averaging (one year before and three months after the major earthquakes). The vertical extent of each box represents the standard deviation of the velocity in the area computed by the method in section 4.2. The horizontal line in each box indicates average velocity $\Delta v$ in each time interval, and the vertical line the center of each time interval. In panel (b), we use only the stations with significant anisotropy $(v_{fast} - v_{slow})/v_{fast} \geq 1\%)$. We select 11 stations from 19 stations and depict the fast shear-wave polarization directions with different symbols or lines. Each symbol is placed at the center of each time period.

mostly on the local geology. The velocity reduction can be due either to the opening and closing of existing fractures, to the creation of new fractures, or to the change in the shear modulus caused by changes in the pore fluid pressure because of shaking-induced compaction (Das, 1993, Figure 4.24).

The relative velocity reduction is smaller than the reduction found by Wu et al. (2009) because of the aver-
We also obtain the polarization directions of the fast shear waves before and after the major earthquakes by averaging over the same time intervals as used in Figure 11a (Figure 11b). The direction of the fast shear-wave polarization does not show a significant change after the earthquakes, hence it seems to be unaffected by these earthquakes. The average standard deviation of the polarization direction of all stations in the Niigata prefecture between 2000 and 2010 is 15 degrees, which represents the accuracy of the fast shear-wave velocity polarization direction.

6.2 Influence of precipitation

We compute the monthly-averaged shear-wave velocities of the north-south polarization (Figure 12a) to investigate a possible seasonal velocity variation related to precipitation. We use only the data in southern Japan (the light shaded area in Figure 1) because that region has a more pronounced seasonal precipitation cycle than northern Japan. Figure 12a illustrates a significant velocity difference between spring/summer and fall/winter. We calculate the average velocities over the stations with the 15% slowest shear-wave velocities in the area because these stations are located at soft-rock sites. We use the stations with the 15% slowest velocities in the area. The horizontal extent of each box shows time interval used for averaging, and the vertical extent the standard deviations of all receivers in the time interval computed by the method in section 4.2. The horizontal line in each box indicates average velocity $\Delta v$ in each time interval, and the vertical line the center of each time interval. (b) Crossplot between monthly precipitation (provide by JMA) and the average velocity $\Delta v$ with error bars. (c) Crossplot between monthly precipitation and the average velocity $\Delta v$ with error bars using the stations with the 85% fastest velocities in the area.

Figure 12. Seasonal dependence of shear-wave velocity. (a) Variation of the average monthly-velocities stacked over the period 2000 through 2010 in southern Japan (the light gray area in Figure 1). We use the stations with the 15% slowest velocities in the area. The horizontal extent of each box shows time interval used for averaging, and the vertical extent the standard deviations of all receivers in the time interval computed by the method in section 4.2. The horizontal line in each box indicates average velocity $\Delta v$ in each time interval, and the vertical line the center of each time interval. (b) Crossplot between monthly precipitation (provide by JMA) and the average velocity $\Delta v$ with error bars. (c) Crossplot between monthly precipitation and the average velocity $\Delta v$ with error bars using the stations with the 85% fastest velocities in the area.
sites and are therefore influenced more by precipitation than the station at hard-rock sites. We compare the monthly-averaged velocities with the monthly average of precipitation (observed by the Japan Meteorological Agency (JMA)) computed from precipitation records over 30 years (Figure 12b). Note the negative correlation between the shear-wave velocity and precipitation (i.e., when it rains, the velocity decreases), which is consistent with the findings of Sens-Schönfelder and Wegler (2006).

For comparison, for the stations with the 85% fastest shear-wave velocities in the area (Figure 12c) the shear-wave velocity does not vary with precipitation. The cause of the velocity reduction is the decreased effective stress of the soil due to the infiltration of water that increases the pore pressure (Das, 1993, Section 4.19; Chapman and Godin, 2001; Snieder and van den Beukel, 2004). We assume that for soft-rock sites most of the velocity change is caused by the effective stress change because Snieder and van den Beukel (2004) show that the relative density change with pore pressure is much smaller than the relative change in the shear modulus.

6.3 Shear-wave splitting and the direction of the plate motion

Using shear-wave splitting analysis, we determine fast shear-wave polarization directions of every station (illustrated by the black arrows in Figure 13a). These directions are averaged over all years from 2000 to 2010 because the temporal changes in the direction are small (see Figure 11b). We plot the directions of all stations which have an anisotropy parameter \((v_{\text{fast}} - v_{\text{slow}})/v_{\text{fast}} \geq 1\%) because the uncertainty in the direction of the fast shear polarization is large when the anisotropy is small. In Figure 13a, we also plot the direction of the plate motion at each station (the gray arrows), estimated from GPS data (Sagiya et al., 2000). Each arrow is normalized to the same length.

In Figure 13b, we plot only the stations which have an anisotropy parameter larger than 1% and a north-south polarization shear-wave velocity faster than 600 m/s; these stations are located on hard-rock sites. The average absolute angle between the directions of fast shear polarization and the plate motion in these stations is 16 degrees, and this average angle of the stations which have a shear-wave velocity less than 600 m/s is 36 degrees. Therefore, the fast shear-wave polarization on hard-rock sites correlates more strongly with the direc-
6. Conclusions

We present a crossplot of the directions of the fast shear-wave polarization and the plate motion for stations all over Japan (Figure 14), where we only used the stations which have an anisotropy parameter greater than 1% and a shear-wave velocity faster than 600 m/s. The red area in Figure 14 indicates that for most stations the direction of the plate motion is between 90 and 140 degrees, and that this direction correlates with the polarization direction of the fast shear wave. The near-surface stress directions on hard-rock sites is presumably related to the plate motion because the stress field related to the plate motion changes the properties of fractures. Note that the used shear waves sample the shallow subsurface (down to about several hundreds of meters). It is remarkable that the shear-wave velocities in the near surface at hard-rock sites correlate with tectonic process (plate motion) that extends several tens of kilometers into the subsurface.

8 Acknowledgments

We are grateful to NIED for providing us with the KIK-net data and to JMA for precipitation data. We thank Diane Witters for her professional help in preparing this manuscript. We are grateful to Kaoru Sawazaki for suggestions, corrections, and discussions.

References

Li, Y.-G., J. E. Vidale, K. Aki, F. Xu, and T. Burdette, 1998, Evidence of shallow fault zone strengthening...
Near-surface S-wave velocities in Japan

...after the 1992 M7.5 Landers, California, earthquake: Science, 279, 217–219.


Miyazawa, M., R. Snieder, and A. Venkataraman, 2008, Application of seismic interferometry to extract P- and S-wave propagation and observation of shear-wave splitting from noise data at Cold Lake, Alberta, Canada: Geophysics, 73, D35–D40.


Tonegawa, T., K. Nishida, T. Watanabe, and K. Sh-...
Seismic interferometry of teleseismic S-wave coda retrieval of body waves: an application to the Philippine Sea slab underneath the Japanese Islands: Geophys. J. Int., 178, 1574–1586.


**APPENDIX A: 1D SEISMIC INTERFEROMETRY**

We explain in this appendix why deconvolution interferometry is suitable for this study. Figure A1 illustrates the model of interferometry using a KiK-net station and an earthquake. The incoming wavefield \( S(r_b, s, \omega) \), propagating from source \( s \) to receiver \( r_b \), is given by \( G(r_b, s, \omega)W(s, \omega) \), where \( G \) is the Green’s function including any unknown complex effect of wave propagation such as scattering and attenuation, and \( W \) the source signature in the frequency domain. Assuming that the subsurface is homogeneous between the receivers, the received wavefield at surface receiver \( r_s \) is

\[
2S(r_b, s, \omega)e^{j2kz_b e^{-2\gamma z_b}},
\]

where \( \gamma \) is the attenuation coefficient and \( k \) the wave number. Because of the free surface, the amplitude of the wavefield at the surface is multiplied by a factor 2. We assume that there are no multiples between the two receivers. The reflected wavefield from the surface at the borehole receiver \( r_b \) is \( S(r_b, s, \omega)e^{2j2kz_b e^{-2\gamma z_b}} \), and the total wavefield at \( r_b \) is \( S(r_b, s, \omega) + S(r_b, s, \omega)e^{2j2kz_b e^{-2\gamma z_b}} \). Applying cross-correlation interferometry to these wavefields yields

\[
S(r_b, s, \omega) = 2S(r_b, s, \omega)e^{j2kz_b e^{-2\gamma z_b}}
\]

\[
= 2|S(r_b, s, \omega)|^2 e^{j2kz_b e^{-2\gamma z_b}}
\]

\[
≈ 2|S(r_b, s, \omega)|^2 e^{j2kz_b e^{-\gamma z_b}}
\]

Figure A1. Geometry of an earthquake and a KiK-net station, where \( r_s \) is the surface receiver (black triangle), \( r_b \) the borehole receiver (white triangle), and \( s \) the epicenter of earthquake (gray star).

\[
\begin{align*}
\text{Figure A1. Geometry of an earthquake and a KiK-net station.} & \\
& \text{where } r_s \text{ is the surface receiver (black triangle), } r_b \text{ the } \\
& \text{borehole receiver (white triangle), and } s \text{ the epicenter of } \\
& \text{earthquake (gray star).}
\end{align*}
\]
in the horizontal direction. In our study, the virtual source in the borehole has the polarization of the incident wave. The two horizontal components of the virtual source therefore are not independent, so that Alford rotation cannot be applied.

The angle of the fast and slow shear-wave polarization directions is 90 degrees because we assume the incoming wave is a plane wave (Cramin, 1985). A wavefield with polarization \( \hat{p} \), which is a unit vector, can be expressed in the polarization of the fast and slow shear waves:

\[
p = \hat{p}_f \cos \phi + \hat{p}_s \sin \phi, \tag{B1}
\]

where \( \phi \) is the polarization angle in the arbitrary wavefield relative to the direction of the fast shear-wave polarization, and \( \hat{p}_f \) and \( \hat{p}_s \) are the unit vectors of the fast and slow velocity wavefields, respectively (see Figure B1).

The incoming wavefield \( u_b \) at the borehole receiver is

\[
u_b = S(t)\hat{p} = S(t)\hat{p}_f \cos \phi + S(t)\hat{p}_s \sin \phi, \tag{B2}
\]

where \( S(t) \) is the incoming wavefield, and \( t \) the time. The wavefield at surface receiver \( u_s \) is given by

\[
u_s = S \left( t - \frac{z_b}{v_f} \right) \hat{p}_f \cos \phi + S \left( t - \frac{z_b}{v_s} \right) \hat{p}_s \sin \phi, \tag{B3}
\]

where \( v_f \) is the fast velocity, \( v_s \) the slow velocity, and \( z_b \) the distance between the top and bottom receivers.

The component of \( u_s \) along \( \hat{p} \) is

\[
u_s(\phi) = (\hat{p} \cdot u_s) = S \left( t - \frac{z_b}{v_f} \right) (\hat{p} \cdot \hat{p}_f) \cos \phi + S \left( t - \frac{z_b}{v_s} \right) (\hat{p} \cdot \hat{p}_s) \sin \phi
\]

\[
= S \left( t - \frac{z_b}{v_f} \right) \cos^2 \phi + S \left( t - \frac{z_b}{v_s} \right) \sin^2 \phi. \tag{B4}
\]

We express \( v_f \) and \( v_s \) in the average velocity \( v_0 \) and the difference \( \delta \):

\[
\frac{1}{v_f} = \frac{1}{v_0} - \delta, \quad \frac{1}{v_s} = \frac{1}{v_0} + \delta, \tag{B5}
\]

and assume the splitting time \( z_b \delta \) is much smaller than the period. We insert expression B5 into equation B4 and expand using first-order Taylor expansion in \( \delta \):

\[\text{Figure B1.} \text{ Projection of fast and slow velocity directions, where } \hat{p}_f \text{ is the fast polarization direction, } \hat{p}_s \text{ the slow polarization direction, } \hat{p} \text{ an arbitrary direction, and } \phi \text{ the angle between the fast direction and arbitrary direction. } \hat{p}, \hat{p}_f, \text{ and } \hat{p}_s \text{ are unit vectors. Dashed arrows show the projection, which is shown in equation B1.}\]
where $S'$ is the time derivative of $S$. Thus, using Taylor expansion, we obtain the velocity for a shear wave with the polarization of equation B1:

$$v(\phi) = v_0 \frac{1}{1 - v_0 \delta \cos 2\phi},$$  \hspace{1cm} (B7)

or to first order in $v_0\delta$:

$$v(\phi) = v_0(1 + v_0 \delta \cos 2\phi).$$  \hspace{1cm} (B8)

**APPENDIX C: FOURIER COEFFICIENTS**

We describe the meaning of $v_0$, $v_1$, and $v_2$ in equation 5. Expression B8 gives the velocity for a polarization $\phi$ relative to the polarization of the fast shear wave. When the azimuth of the fast shear-wave polarization is given by $\psi$, the angle $\phi$ in equation B8 must be changed into $\phi \rightarrow \phi - \psi$. Denoting $v_0^2\delta$ by $V$, these changes turn equation B8 into

$$v(\phi) = v_0 + V \cos 2(\phi - \psi).$$  \hspace{1cm} (C1)

This can also be written as

$$v(\phi) = v_0 + v_1 \cos 2\phi + v_2 \sin 2\phi;$$  \hspace{1cm} (C2)

with

$$v_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} v(\phi) d\phi, \quad v_1 = \frac{1}{\pi} \int_{-\pi}^{\pi} v(\phi) \cos 2\phi d\phi,$$

$$v_2 = \frac{1}{\pi} \int_{-\pi}^{\pi} v(\phi) \sin 2\phi d\phi,$$

$$V = \sqrt{v_1^2 + v_2^2}, \quad \psi = \frac{1}{2} \arctan \left( \frac{v_2}{v_1} \right).$$

In expression C2, $v_0$, $v_1$, and $v_2$ are the Fourier coefficients of the velocity $v(\phi)$. Because $v_0$ does not depend on $\phi$, $v_0$ represents the isotropic velocity. Expression C1 shows that $v(\psi)$ and $v(\psi + \pi/2)$ are the fastest and slowest velocities, respectively. Therefore, $V$ is the anisotropic velocity and the angle between the fast and slow polarization directions is 90 degrees, which corresponds to shear-wave splitting of a plane wave (Crampin, 1985).
Near-surface weakening in Japan after the 2011 Tohoku-Oki earthquake

Nori Nakata & Roel Snieder
Center for Wave Phenomena, Geophysics Department, Colorado School of Mines, Golden, Colorado 80401

ABSTRACT
The magnitude (M_W) 9.0 Tohoku-Oki earthquake on 11 March 2011 was one of the largest in recent history. Ground motion caused by the seismicity around the time of the main shock was recorded by KiK-net, the strong-motion network that covers most of Japan. By deconvolving waveforms generated by earthquakes that are recorded at the surface and in a borehole at KiK-net station FKSH18, we detect a reduction of shear-wave velocity in the upper 100 m of about 10%, and a subsequent healing that varies logarithmically with time. Using all available borehole and surface records of more than 300 earthquakes that occurred between 1 January 2011 and 26 May 2011, we observe a reduction in the shear-wave velocity of about 5% in the upper few hundred meters after the Tohoku-Oki earthquake throughout northeastern Japan. The area of the velocity reduction is about 1,200 km wide, which is much wider than earlier studies reporting velocity reductions following other larger earthquakes. The reduction of the shear-wave velocity is an indication that the shear modulus, and hence the shear strength, is reduced over a large part of Japan.

Key words: seismic interferometry, Tohoku-Oki earthquake, shear-wave velocity, near surface, KiK-net

1 INTRODUCTION
The Tohoku-Oki earthquake (M_W 9.0) of 11 March 2011 is one of the largest earthquakes in recent times. The subduction of the Pacific Plate at a velocity of 8-8.5 cm/year (DeMets et al., 2010) has resulted in many M_W 7+ earthquakes (Miyazawa and Mori, 2009). Before and after the Tohoku-Oki earthquake, many smaller earthquakes occurred. We use ground motion excited by seismicity recorded by KiK-net (the strong-motion network operated by the National Research Institute for Earth Science and Disaster Prevention (NIED)) to estimate time-lapse changes of the shear-wave velocities in the shallow subsurface throughout northeastern Japan after the Tohoku-Oki earthquake.

To measure shear-wave velocities, we use seismic interferometry, developed over the last 10 years (Claerbout, 1968; Trampert et al., 1993; Lobkis and Weaver, 2001; Roux and Fink, 2003; Schuster et al., 2004; Wapenaar, 2004; Bakulin and Calvert, 2006; Snieder et al., 2006) to determine the arrival time of waves that propagate between two sensors. This technique has been applied to earthquake data in various ways, such as measuring shear-wave velocity (e.g., Snieder and Şafak, 2006; Sawazaki et al., 2009) and estimating deep subsurface structure (e.g., Tonegawa et al., 2009; Ruigrok et al., 2010).

In this paper, we present the time-lapse change of the near-surface shear-wave velocity throughout the east half of Japan after the Tohoku-Oki earthquake. First, we introduce the use of KiK-net data based on seismic interferometry and time interpolation. Then we show the waveforms of one KiK-net station retrieved by seismic interferometry. Finally, we present a shear-wave velocity-change map throughout northeastern Japan.

2 KIK-NET
About 700 KiK-net stations are distributed across Japan (Okada et al., 2004). Each station has a borehole with three-component strong-motion seismographs at the bottom and top of the borehole. The sampling interval of KiK-net stations is 0.01 s.

We use all available KiK-net stations and seismicity from 1 January 2011 to 26 May 2011. The depths of borehole seismometers are between 100 m and 337 m (91% of the seismometers are at a depth less than
210 m). Magnitude of seismicity is between 2.8 and 9.0. The observed record, as used for seismic interferometry, ranges from 60 s to 300 s depending on the earthquake.

All the used events are at a depth greater than 7 km, which is a relatively large depth compared to the depth of the boreholes. The velocity in the near surface is much slower than it is at greater depths. Because of the depth of events and slow velocities in the near surface, the waves that travel between the sensors at each station propagate in the near-vertical direction. Hence we assume the incoming waves at the receivers are locally near-vertical plane waves. In this study, we use only the north-south horizontal component. Before the data processing, we apply a bandpass filter from 1 to 13 Hz for all earthquake data.

3 COMPUTING METHODS
3.1 Deconvolution interferometry
Seismic interferometry is a technique to obtain the Green’s function that accounts for wave propagation between two stations (Claerbout, 1968; Lobkis and Weaver, 2001; Roux and Fink, 2003; Wapenaar, 2004; Bakulin and Calvert, 2006; Snieder et al., 2006). Although the widest applied algorithm in seismic interferometry is based on cross correlation (e.g., Claerbout, 1968; Wapenaar, 2004; Bakulin and Calvert, 2004; Schuster et al., 2004), we use the algorithm based on deconvolution (e.g., Trampert et al., 1993; Snieder and Şafak, 2006; Vasconcelos and Snieder, 2008). In deconvolution interferometry, we can suppress the complicated imprint of the structure (e.g., attenuation and scattering) incurred as the waves travel from the hypocenter to the borehole seismogram (Snieder et al., 2009). We denote the wavefield excited by an earthquake at location \( s \) that strikes the borehole receiver at location \( r_b \) by \( u(r_b, s, \omega) \), and the wavefield recorded at the surface receiver at location \( r_s \) by \( u(r_s, s, \omega) \) in the frequency domain. The deconvolved waveform is given by

\[
D(\omega) = \frac{u(r_s, s, \omega)}{u(r_b, s, \omega)} \approx \frac{u(r_s, s, \omega) u^*(r_b, s, \omega)}{\left| u(r_b, s, \omega) \right|^2 + \epsilon},
\]

(1)

where \( * \) is the complex conjugate and \( \epsilon \) the regularization parameter that stabilized the deconvolution (Snieder and Şafak, 2006; Mehta et al., 2007). \( D(\omega) \) is the frequency-domain waveform that propagates from the borehole sensor to the surface sensor (Snieder and Şafak, 2006; Mehta et al., 2007; Sawazaki et al., 2009; Yamada et al., 2010). We choose \( \epsilon \) to be 1% of the average power spectrum of the wavefield at the borehole receiver because we find experimentally this is the smallest value of the regularization parameter that produces stable deconvolved wavefields.

3.2 Enhancing time resolution
Because the sampling time of KiK-net seismometers is larger than the changes in the travel time that we seek to measure, we interpolate the deconvolved waveforms and enhance the time resolution. We estimate the arrival time by selecting the three adjacent samples with the largest amplitude and quadratically interpolate between these points. We use the time of the maximum amplitude of the parabola thus obtained as the arrival time of the deconvolved wave. This makes it possible to measure the arrival time with a resolution better than the sampling time. We refer to this procedure as quadratic interpolation.

3.3 Estimating the angle of incidence
We compute the angle of the incoming wave at the borehole receiver by using one-dimensional ray tracing to confirm whether the wave propagating between the borehole and surface sensors propagates vertically. We use the velocity model of Nakajima et al. (2001) to determine the ray parameter \( p \) of the ray that connects each earthquake with the borehole sensor. The angle of incidence \( \theta \) of the wave that propagates between the borehole and surface seismometers is then given by \( \cos \theta = \sqrt{1 - p^2 v^2} \), where \( v \) is the average shear-wave velocity between these sensors as determined in this study. A bias in the velocity estimation due to non-vertical propagation depends on the deviation from \( \cos \theta \) from its value for vertical incidence, \( \cos 90^\circ = 1 \).

4 DETERMINING SHEAR-WAVE VELOCITIES THROUGHOUT NORTHEASTERN JAPAN
Figure 1a shows deconvolved waveforms of earthquakes between 1 January 2011 and 26 May 2011 for KiK-net station FKSH18 in the Fukushima prefecture at a distance of about 200 km from the epicenter of the Tohoku-Oki earthquake; Figure 1b shows the epicenters of the earthquakes that occurred during the periods before and after the event. The arrival times obtained by quadratic interpolation are shown with circles in Figure 1a. The average of \( \cos \theta \) (see section 3.3) over the events between 1 January 2011 and 10 March 2011 is 0.984, while between 12 March 2011 to 26 May 2011 this average is equal to 0.980. This implies that the bias in the estimated shear-wave velocity is only about 2%, but this bias is virtually identical in the periods before and after the Tohoku-Oki earthquake. Hence changes in the pattern of seismicity before and after the main shock are not responsible of the change in the shear-wave velocity that we present. To enhance the data quality, we discard some data which has a low signal-to-noise ratio based on a visual inspection.

The travel time measured during the main shock
Near-surface weakening in Japan

Figure 1. (a) Deconvolved waveforms of individual earthquakes from 1 January 2011 to 26 May 2011 at station FKSH18. This station recorded 25 earthquakes from 1 January 2011 to 10 March 2011 (black curves), the Tohoku-Oki earthquake of 11 March 2011 (magenta thick curve), 5 other earthquakes on 11 March 2011 (red curves), and 96 earthquakes from 12 March 2011 to 26 May 2011 (blue curves). Circles, marked by the same color as the waveforms (blue replaces cyan), represent the interpolated arrival times of waves. The waveforms are ordered by the origin times of earthquakes in the vertical axis. (b) Epicenters of two time intervals: 1 January 2011 to 10 March 2011 and 12 March 2011 to 26 May 2011. The size of each circle indicates the magnitude of each earthquake and the color denotes the depth. The white triangle points to the location of station FKSH18.

of the Tohoku-Oki earthquake (the large magenta circle in Figure 1a) is significantly later than that from the other earthquakes. This indicates a reduction of the shear-wave velocity of about 22% during the shaking caused by the Tohoku-Oki event. Note also the delay of the waves in the early aftershocks indicated in red in Figure 1a. The delay of the waveforms after the Tohoku-Oki earthquake relative to the waveforms recorded before the event indicates that the shear waves propagate with a reduced shear-wave velocity after the Tohoku-Oki earthquake (Figure 1a).

Figure 2 depicts the travel-time change during the shaking caused by the Tohoku-Oki earthquake by applying short-time moving-window seismic interferometry to the seismogram, in which we deconvolve 20-s time windowed borehole and surface records at station FKSH18. Since the time window moves with 10-s intervals, the windows have a 10-s overlap. The main delay occurs at 30-40 s, and it is increasing while the shaking increases. After the strongest shaking (at 130 s), the travel times recover and are fairly constant. Note that the delay, as well as the shear-wave velocity reduction, remains nonzero after 200 s compared to its values between 0-20 s. The velocity reduction at the time of strong shaking is likely to be influenced by several physical mechanisms including incipient liquefaction.

We compute the shear-wave velocity as a function of time from the interpolated travel times (the circles in Figure 1) using the known depth of the borehole. Figure 3 shows the shear-wave velocity estimated from each earthquake at station FSKH18. According to Figure 3, the velocity is reduced by almost 10% on the day after the Tohoku-Oki earthquake and the velocity recovers with about 5% in the 2 months after the earthquake. As shown by the orange curve in Figure 3, the shear-wave velocity after the Tohoku-Oki earthquake recovers logarithmically with time (Dieterich, 1972; Vidale and Li, 2003; Schaff and Beroza, 2004), $v_s(t) = a \ln(t - t_0) + b$, where $t_0$ is the origin time of the Tohoku-Oki event, and $t$ is time measured in days. We determine the parameters $a$ and $b$ by a linear least-squares fit of the data points shown by the red and blue dots in Figure 3. We exclude the data point of the Tohoku-Oki earthquake (the large magenta dot in Figure 3) in the esti-
Figure 2. Short-time moving-window seismic interferometry of the ground motion caused by the Tohoku-Oki earthquake. (a) The earthquake record observed at the north-south horizontal component borehole seismometer of station FKSH18. Gray bars indicate the 20-s time windows for seismic interferometry with 10-s overlap. Black circles are the center of each window. (b) Deconvolved waveforms of each time window. Each waveform is aligned with the center time of the employed time window. Black circles illustrate the interpolated arrival times.

mation of the orange recovery curve in Figure 3 because the anomalously low velocity during the shaking by the Tohoku-Oki event may be caused by a complex physical mechanism mentioned above.

We compute the average of the deconvolved waveforms for station FKSH18 over the periods 1 January - 10 March (before) and 12 March - 26 May (after) in 2011 of Figure 1. These average waveforms are shown by the solid lines in Figure 4. The shapes of the average deconvolved waveforms before and after the Tohoku-Oki earthquake are similar, but the average waveform after the Tohoku-Oki earthquake is delayed. We also determine the average shear-wave velocities before and after the Tohoku-Oki earthquake from the interpolated travel times (the circles in Figure 4). The average velocity in the time interval before the Tohoku-Oki earthquake is 665 ± 7 m/s, and after the event it is 625 ± 14 m/s, hence the average velocity reduction is about 6%. The uncertainty of the velocities is determined from the standard deviations of the travel times over all events in each time interval.

It has been documented that the shear-wave velocity in the near surface may exhibit seasonal changes associated with changes in precipitation (Sens-Schönfelder and Wegler, 2006). In order to investigate the influence of seasonal changes, we compute the mean shear-wave velocities in the periods 1 January - 10 March and 12 March - 26 May averaged over all years from 2000 to 2010. The corresponding waveforms are shown by the dashed lines in Figure 4. The mean velocity over the period 1 January - 10 March averaged from 2000-2010 is 664 ± 6 m/s, and the mean velocity for the interval 12 March - 26 May is 661 ± 6 m/s. The difference between these values is statistically not significant, and it is much smaller than the measured velocity change associated with the Tohoku-Oki earthquake.

We average the deconvolved waves at each KiK-net station over earthquakes recorded in the time intervals before (from 1 January 2011 to 10 March 2011) and after (from 12 March 2011 to 26 May 2011) the Tohoku-Oki event to determine the arrival times of the average deconvolved waveforms at each KiK-net station that are the travel time of the shear wave that propagates between the seismometers in the borehole and at the surface of each station. These times thus constrain the near-surface shear-wave velocity between the seismometers. We convert this travel time to the shear-wave velocity in the near-surface at each station, and following spatial interpolation (Lawson, 1984) of the velocities between stations, we obtain near-surface shear-wave velocity maps before (the upper-left map in Figure 5) and after (the middle map in Figure 5) the Tohoku-Oki earthquake. In order to reduce the uncertainty in the velocity estimates, we use only stations that recorded more than 3 earthquakes during both time intervals. The average cosine of θ is greater than 0.975 but in the west side of the area cosine θ ≈ 0.94 – 0.96. These values are fairly constant in the time periods before and after the Tohoku-Oki earthquake. We use recorded data from 83

Figure 3. Shear-wave velocity variations in the upper 100 m at station FKSH18. By using the arrival times of waves (the circles in Figure 1), we compute the velocity variations from 1 January 2011 to 26 May 2011. The color of each dot is the same as in Figure 1. Black vertical line indicates the origin time of the Tohoku-Oki earthquake. Orange line depicts the average velocity (before the Tohoku-Oki earthquake) and the logarithm curve determined by least-squares fitting of the velocity after the earthquake. We do not include the Tohoku-Oki earthquake data point (magenta dot) in the data fit.
Near-surface weakening in Japan

Figure 4. Averaged waveforms of Figure 1 before (from 1 January 2011 to 10 March 2011; black solid curve) and after (from 12 March 2011 to 26 May 2011; blue solid curve) the Tohoku-Oki earthquake at station FKSH18. Circles denote the interpolated arrival times of averaged waves. Black and blue dashed curves represent the averaged waveforms from 1 January to 10 March and from 12 March to 26 May over 11 years (from 2000 to 2010), respectively.

and 219 earthquakes, respectively, to create shear-wave velocity maps for the time intervals before and after the Tohoku-Oki earthquake. By subtracting the velocity measured before the main event from the velocity measured after the event, we obtain the map of the relative velocity change before and after the Tohoku-Oki earthquake shown in the lower-right map of Figure 5.

5 DISCUSSION AND CONCLUSIONS

It is known that large earthquakes can reduce seismic velocities close to the epicenter (e.g., Li et al., 1998; Vidale and Li, 2003; Schaff and Beroza, 2004; Wegler and Sens-Schönfelder, 2007; Brenguier et al., 2008; Sawazaki et al., 2009; Yamada et al., 2010). As shown in Figure 5, the shear-wave velocity was reduced by about 5% after the Tohoku-Oki earthquake over an area in northeastern Japan about 1,200 km wide, which is much larger than the region of velocity reduction after the earthquakes reported in earlier studies. We also measured the mean shear-wave velocity reduction in these time intervals over the period from 2000 to 2010 of the whole area shown in the maps in Figure 5. The seasonal change in the shear-wave velocity is only 0.2%, which is much smaller than the velocity reduction observed following the Tohoku-Oki earthquake (see the lower-right map of Figure 5). We conclude that the shear-wave velocity reduction in Figure 5 is caused by the Tohoku-Oki earthquake. The area with reduced shear-wave velocity is delimited on the western side by the Median Tectonic Line (MTL) and the Itoigawa-Shizuoka Tectonic Line (ISTL) (the dashed black lines on the lower-right map in Figure 5). Because the number of recorded earthquakes at the west side of these tectonic lines is small, between 3 and 5, and the average of \( \cos \theta \) is relatively low (around 0.94-0.96), the velocities in the western part are less reliable than those in other regions. The velocity reduction of Figure 5 does not correlate with the coseismic or postseismic displacements of the Tohoku-Oki earthquake (Ozawa et al., 2011) because the velocity reduction is also influenced by variations in local geology.

With seismic interferometry, we extract the waves that propagate between the borehole and surface seismometers at KiK-net stations, and find a significant
reduction of the near-surface shear-wave velocity after the Tohoku-Oki earthquake that recovers logarithmically with time. By applying this analysis to all available seismograms, we detect a reduction of the shear-wave velocity in the upper few hundred meters throughout northeastern Japan. The shear-wave velocity is related to the shear modulus; hence the reduction of the shear-wave velocity over northeastern Japan implies that the Tohoku-Oki earthquake reduced the shear strength of the near surface throughout northeastern Japan.

ACKNOWLEDGMENTS

We thank NIED for providing us with the KiK-net data. We appreciate David Schaff for suggestions, corrections, and discussions.

REFERENCES


Trampert, J., M. Cara, and M. Frogneux, 1993, SH
propagator matrix and $Q_s$ estimates from borehole- and surface- recorded earthquake data: Geophys. J. Int., 112, 290–299.
Time-lapse change in anisotropy in Japan’s near surface caused by the 2011 Tohoku-Oki earthquake

Nori Nakata & Roel Snieder
Center for Wave Phenomena, Geophysics Department, Colorado School of Mines, Golden, Colorado 80401

ABSTRACT
We apply seismic interferometry to strong-motion records to detect the change in anisotropy caused by the MW 9.0 Tohoku-Oki earthquake on 11 March 2011. It is known that large earthquakes change fluid condition in cracks, and thereby the polarization-anisotropy condition. The Tohoku-Oki earthquake increased the difference between fast and slow shear-wave velocities arising from shear-wave splitting in most parts of northeastern Japan, but it did not significantly change polarization directions of the fast shear waves in the upper few hundred meters. The difference between fast and slow velocities increased gradually for more than one month after the main shock, whereas the decrease in velocity occurred suddenly after the event. Through monitoring of anisotropy and shear-wave velocity, we find that the changes in anisotropy and velocity recover with time. The change in anisotropy are correlated with the change in velocity in the northeastern Honshu. Also, the change in the largest principal stress direction weakly correlates with that in anisotropy.

1 INTRODUCTION
The change in near-surface shear-wave velocity caused by the MW 9.0 Tohoku-Oki earthquake on 11 March 2011 is noted by Nakata and Snieder (2011). The earthquake, among the largest in recent history, resulted in a reduction in the near-surface velocity averaged over two months following the earthquake of about 5% throughout northeastern Japan (a region 1,200-km wide). In this study, we estimate the change in near-surface polarization anisotropy by applying seismic interferometry to seismograms recorded by KiK-net, a strong-motion recording network operated by the National Research Institute for Earth Science and Disaster Prevention (NIED).

At a depth greater than a few kilometers, the polarization anisotropy in northeastern half of Japan is related to the mantle wedge; the fast shear-wave polarization directions on the fore-arc and back-arc sides are approximately along the north-south and east-west directions, respectively (Okada et al., 1995; Nakajima and Hasegawa, 2004). Local geology, however, alters the shear-wave polarization in the near surface, a few hundred meters deep, and even at hard-rock sites the fast shear-wave polarization directions on the fore-arc side do not agree with the deep anisotropy (Nakata and Snieder, 2012).

Conventionally, shear-wave splitting is estimated by the cross-correlation method (e.g., Fukao, 1984). One can estimate the fast and slow shear-wave polarization directions and the delay time between the fast and slow shear waves, which are a mean value along a ray path. Moreover, using a cluster of earthquakes, one can estimate the vertical variation of anisotropy (e.g., Okada et al., 1995). The polarization directions and the delay time are changed by large earthquakes (e.g., Cochran et al., 2006). Some previous studies have found the delay time to increase after intermediate or large earthquakes, but the directions do not change (e.g., Saiga et al., 2003; Liu et al., 2004) because the delay time is more sensitive to change in stress than the polarization direction is (Peacock et al., 1988). Since changes in delay times have been observed prior to major earthquakes (e.g., Peacock et al., 1988; Crampin et al., 1990; Crampin and Gao, 2005), monitoring the delay time has been proposed as a diagnostic for earthquake prediction (Crampin et al., 1984b).

We present the change in anisotropy based on shear-wave splitting caused by the Tohoku-Oki earthquake inferred from seismic interferometry. First, we show shear-wave splitting of individual earthquake signals at one KiK-net station. Then we compute changes in polarization anisotropy after the main event for all available stations and compare with the changes in shear-wave velocity and the largest principal stress direction caused by the main shock.
Figure 1. Variation in shear-wave velocity and anisotropy coefficient from 1 May 2010 to 31 December 2011 at station FKSH12. The top panel depicts the isotropic velocity of each earthquake (black dot) and its nine-point moving average (blue line). The bottom panel indicates the anisotropy coefficient computed from fast and slow shear-wave velocities (black cross) and its nine-point moving average (blue line). The velocity and the anisotropy coefficient estimated from the main event are illustrated by magenta symbols. Green vertical lines denote the origin time of the event. Red horizontal lines and gray shaded areas are the mean values and the mean values ± the standard deviations of the measurements of all used earthquakes during each period. We show mean values of each period in Table 1.

2 EARTHQUAKE RECORDS AND THE ANALYZING METHOD

2.1 KiK-net

KiK-net, which includes about 700 stations all over Japan, has recorded strong motions continuously since the end of the 1990s (Okada et al., 2004). Each KiK-net station has a borehole a few hundred meters deep and two three-component seismographs, with a 0.01-s sampling interval, at the top and bottom of the borehole. All the earthquakes used in this study are at a depth greater than 7 km, which is large compared to the depth of the boreholes. The velocity in the near surface is much slower than at greater depths. Since we consider events much deeper than the borehole, and because of the slow velocities at the near surface, we assume the waves propagate from the borehole receiver to the surface receiver as plane waves in the vertical direction at each station.

To confirm this assumption, we compute the angle of incidence $\theta$ by employing the procedure proposed by Nakata and Snieder (2012) using ray tracing. All earthquake data used have $\cos \theta > 0.975$, which means the maximum of the estimated velocity bias is 2.5%. The bias is, in practice, much smaller because of averaging over many earthquakes. This inaccuracy is not sensitive to the analysis of shear-wave splitting because $\cos \theta$ is the same for the waves in all polarization directions.

2.2 Seismic interferometry

We apply deconvolution-based seismic interferometry to the seismograms of each station to retrieve the wave that propagates from the borehole receiver to the surface receiver. We deconvolve the seismogram of the borehole sensor from that of the surface sensor in the frequency domain (Snieder and Şafak, 2006):

$$D(\omega) = \frac{u(r_s, s, \omega)}{u(r_b, s, \omega)} \approx \frac{u(r_s, s, \omega)u^*(r_b, s, \omega)}{|u(r_b, s, \omega)|^2 + \epsilon}, \quad (1)$$

where $r_s$ and $r_b$ are the locations of the surface and borehole receivers respectively, $s$ is the position of the hypocenter of an earthquake, $u(r, s, \omega)$ is the waveform recorded at $r$, and $\epsilon$ is a regularization parameter to prevent instability of the deconvolution. In this study, we use $\epsilon$ as 1% of the average of power spectrum $|u(r_b, s, \omega)|^2$, which is the smallest value necessary to obtain stable deconvolved waveforms (determined by a visual inspection), and we apply a bandpass filter from 1 to 13 Hz after computing the deconvolved waveforms.

To determine the fast and slow polarization directions caused by shear-wave splitting, we follow the
3 CHANGE IN ANISOTROPY CAUSED BY THE TOHOKU-OKI EARTHQUAKE

We first present earthquake records of KiK-net station FKSH12, which is in the Fukushima prefecture (220 km west-southwest from the main-shock epicenter). Earthquakes used in this study were recorded from 1 May 2010 to 31 December 2011, the magnitude range is confined from 3.0 to 9.0. We compute the isotropic shear-wave velocities and the anisotropy coefficients, \((v_{\text{fast}} - v_{\text{slow}})/t_{\text{fast}}\) (where \(v_{\text{fast}}\) and \(v_{\text{slow}}\) are the fast and slow velocities, respectively), of each earthquake (Figure 1). Dividing the anisotropy coefficient by the slow velocity gives the travel-distance-normalized delay time \((t_{\text{slow}} - t_{\text{fast}})/D\) (where \(t_{\text{slow}}\) and \(t_{\text{fast}}\) are the travel times of slow and fast shear waves respectively, and \(D\) the length of the ray path), which has been used in previous studies (e.g., Saiga et al., 2003; Liu et al., 2004). This is the difference between the average slownesses of the fast and slow waves.

In Table 1, we show the mean values of isotropic shear-wave velocity and anisotropy coefficients during the periods of 1 May 2010–10 March 2011, 12 March 2011–26 May 2011, and 27 May 2011–31 December 2011 (the red lines in Figure 1). Based on Student’s t-test (e.g., Bulmer, 1979), the mean velocities and mean anisotropy coefficients are significantly different between each pair of periods (greater than 99.7% confidence). After the main shock, the shear-wave velocity decreases and the anisotropy coefficient increases, and these changes recover with time (mean velocity: 770–723–743 m/s and mean anisotropy coefficient: 7.8–12.5–10.8%). Nakata and Snieder (2011) discuss the change in the shear-wave velocity caused by the Tohoku-Oki earthquake. Because the fluid condition in cracks is one major cause of anisotropy (Crampin et al., 1984a), large and intermediate earthquakes, which induce the stress change to open or close cracks (Nur and Simmons, 1969) and extend cracks (Atkinson, 1984), can change the anisotropy coefficient.

As shown by the moving average of the anisotropy coefficient (the blue line in the bottom panel of Figure 1), the anisotropy coefficient continues to increase for more than one month after the main shock, which we might attribute to several large aftershocks during that period. Saiga et al. (2003) observe similar phenomena. In contrast, the velocity decreases suddenly after the main shock (see the blue line in the top panel of Figure 1).

The moving average of the anisotropy coefficient decreases before the main shock (the blue line in the bottom panel of Figure 1), but it is not as significant as the changes between each pair of periods. Some studies report changes in anisotropy before large earthquakes (e.g., Crampin et al., 1990; Crampin and Gao, 2005). We, however, need to consider the influence of inter-

---

**Table 1.** Mean values of isotropic shear-wave velocity and anisotropy coefficient at station FKSH12 over all used earthquakes in each period (the red lines in Figure 1). The range of each value is the 95% confidence interval of the mean.

<table>
<thead>
<tr>
<th>Period</th>
<th>Number of earthquakes used</th>
<th>Isotropic velocity (m/s)</th>
<th>Anisotropy coefficient (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 May 2010–10 March 2011</td>
<td>61</td>
<td>763–777</td>
<td>7.2–8.4</td>
</tr>
</tbody>
</table>
mediate earthquakes that occur before the main shock, such events may change the anisotropy as well. For example, the M6.2 earthquake on 13 June 2010 (a distance of 100 km and a depth of 40 km from station FKSH12) and the M5.7 earthquake on 29 September 2010 (a distance of 40 km and a depth of 8 km from the station) might both have been sources of elevated the anisotropy coefficient. The absence of such intermediate events in the nine weeks before the main earthquake near the station could have caused the observed reduction of the anisotropy coefficient in that period.

We compute the fast and slow polarization directions and estimate the anisotropy coefficient for all available stations throughout Japan for a period before the major earthquake (1 January 2011–10 March 2011) and a period afterward (12 March 2011–26 May 2011) (Figure 2). To reduce uncertainty, we use only stations that have 1) more than three earthquake records during both time intervals, 2) travel times of interferometric waves greater than 0.1 s, 3) anisotropy coefficients greater than 1%, and 4) a standard deviation of velocity measurements smaller than 5%. The average change in the angles of the fast shear-wave polarization directions before and after the main event over all used stations is 17 degrees; this is close to the uncertainty, 15 degrees, computed from data over 11 years (Nakata and Snieder, 2012). We conclude that the fast shear-wave polarization direction does not change significantly as a result of the main shock.

In contrast, the anisotropy coefficient in most parts of northeastern Japan increases after the earthquake. To evaluate the change in the anisotropy coefficient caused by the event, we define the change in anisotropy as \((AC_{after} - AC_{before})/AC_{before}\), where \(AC_{before}\) and

Figure 2. Changes in shear-wave velocity and anisotropy coefficient after the Tohoku-Oki earthquake. Each map has a label at upper-right: anisotropy coefficients before (Before: 1 January 2011 to 10 March 2011) and after (After: 12 March 2011 to 26 May 2011) the Tohoku-Oki earthquake, its change, defined as \((AC_{after} - AC_{before})/AC_{before}\) (Anisotropy change), and the change in the isotropic shear-wave velocity (Velocity change). Dark blue (before) and light blue (after) arrows on the Before, After, and Anisotropy-change maps represent the direction of fast shear-wave polarization. The longitude and latitude pertain to the leftmost map. The dashed black lines show the locations of major tectonic lines (the Median Tectonic Line and the Itoigawa-Shizuoka Tectonic Line) (Ito et al., 1996). Locations and magnitude of the earthquakes from 1 January 2011 to 26 May 2011 are shown as circles and relative to the rightmost map. The size of each circle indicates the magnitude of each earthquake and the color represents its depth. The yellow star denotes the epicenter of the Tohoku-Oki earthquake. The small Japanese map at the top shows four regions for interpretation in Figure 3.
S-WAVE SPLITTING CAUSED BY THE TOHOKU EQ

Figure 3. Crossplot of the changes in shear-wave velocity and anisotropy coefficient in regions illustrated in the small map at the top in Figure 2. Each symbol indicates the data of each station. The numbers in the corners of each panel show the fraction of stations in each quadrant of all stations in each region.

$AC_{after}$ are the anisotropy coefficients before and after the main shock, respectively. The change in the anisotropy coefficient is shown in the second map from right in Figure 2. In Figure 3, we show a crossplot of the changes in the shear-wave velocity and the anisotropy coefficient in four regions defined by the small map in Figure 2. The changes are reasonably well correlated in region II (which is the closest region to the epicenter) and poorly correlated in region I. Different from other regions, most measurements in region IV are in the lower-right quadrant where the velocity increases and the anisotropy coefficient decreases. Region IV is on the west side of the tectonic lines (the Median Tectonic Line and the Itoigawa-Shizuoka Tectonic Line: the black dashed lines in Figure 2), and the geologic age and geomorphological classification both differ across these lines; the west side is an older mountain area, and the east side consists of younger volcanics and sediments (Wakamatsu et al., 2006).

4 COMPARING THE CHANGES IN ANISOTROPY AND STATIC STRESS

The Tohoku-Oki earthquake changed the stress and strain conditions (Hasegawa et al., 2011). Changes in stress and strain induce changes in local permeability and pore pressure (Koizumi et al., 1996), and thereby changes in the anisotropy coefficient (Zatsepin and Crampin, 1997). Fluid-filled microcracks, which cause shear-wave splitting (Zatsepin and Crampin, 1997), usually align with the direction of $in situ$ stress (Crampin, 1978). Changes in stress caused by intermediate and large earthquakes have been studied for decades (e.g., Hanks, 1977; King et al., 1994; Baltay et al., 2010). Saiga et al. (2003) compare at two stations the time delays associated with shear-wave splitting with the change in Coulomb stress, which is an indicator of how close a fault is to failure (e.g., King et al., 1994). Toda et al. (2011) and Yoshida et al. (2012) compute the change in stress in northeastern Japan caused by the Tohoku-Oki earthquake.

We compare the change in the anisotropy coefficient with the change in the largest principal stress direction computed by Yoshida et al. (2012) (Figure 4). In Figure 4b, the largest principal change in the stress direction and the change in the anisotropy coefficient indicate a weak positive correlation except for the areas B, J, K, and L. Note that the change in the largest principal stress direction is only one proxy of the changes in stress, and we cannot explain the change in the anisotropy coefficient in area B from the change in the principal stress direction. A large change in the principal stress direction ($> 20^\circ$) signifies that the stress condition before and after the main shock is significantly changed; therefore a large change in the principal stress direction might induce the large change in the anisotropy coefficient ($> 10\%$) in areas A, C, and G. Likewise, a small change in the principal stress direction ($< 20^\circ$) is coincident with the small change in the anisotropy coefficient ($< 10\%$) in areas D, E, F, H, I, and M.

Areas J and K (the asterisks in Figure 4b) are on the west side of the tectonic lines, and area L is close to these lines; thus the change in stress caused by the main earthquake in the upper few hundred meters (the depth range of the boreholes) might differ on the two sides of the tectonic lines. Kern (1978) found in rock-physics experiments that as the confining pressure increases, velocity increases and anisotropy decreases. We speculate that the increase in the velocity and the decrease in the anisotropy coefficient on the west side of the tectonic lines could be explained by increase in the compressional stress, but since we cannot directly measure the compressional stress in this study, we cannot...
Figure 4. (a) Anisotropy change in Figure 2 with the largest principal stress directions before (red arrows: the same as the red arrows in Figure 3a in Yoshida et al. (2012)) and after (blue arrows: the same as the red arrows in Figure 3b in Yoshida et al. (2012)) the Tohoku-Oki earthquake. The arrows are estimated in each 0.5°-grid area. Black dashed lines indicate the locations of the major tectonic lines. A-M areas denote the interpreted regions in panel (b) as well as Yoshida et al. (2012). (b) Crossplot of the changes in the largest principal stress direction (Yoshida et al., 2012) and the anisotropy coefficient in each area shown in panel (a). The change in the anisotropy coefficient is the mean value for each 0.5° grid. Asterisks indicate areas on the west side of the tectonic lines.

validate this hypothesis. Note that the inversion model in Yoshida et al. (2012) does not include possible differences in compaction and rheology across these lines.

5 CONCLUSION

By applying deconvolution-based seismic interferometry to KiK-net data, we measure changes in anisotropy caused by the Tohoku-Oki earthquake. The anisotropy coefficient increases in most parts of northeastern Japan after the Tohoku-Oki earthquake, but the fast polarization direction does not change. The changes in shear-wave velocity and anisotropy both recover with time. Comparison of the changes in the shear-wave velocity and the anisotropy coefficient shows strong correlation in the eastern half of Honshu have a strong correlation. The changes in the anisotropy coefficient and the largest principal stress direction are weakly correlated. On the west side of the tectonic lines, the increase in velocity and the decrease in anisotropy could be explained by a difference of the change in stress across the tectonic lines.

ACKNOWLEDGMENTS

We are grateful to NIED for providing us with the KiK-net data.

REFERENCES

Crampin, S., and Y. Gao, 2005, Comment on “Systematic analysis of shear-wave splitting in the aftershock zone of the 1999 Chi-Chi, Taiwan, earthquake: Shallow crustal anisotropy and lack of precursory changes,
S-WAVE SPLITTING CAUSED BY THE TOHOKU EQ


Estimation of velocity change using repeating earthquakes with different locations and focal mechanisms

Chiaemerem O. Kanu and Roel Snieder
Center for Wave Phenomena, Colorado School of Mines, Golden CO 80401

ABSTRACT
Codas of repeating earthquakes carry information about the time-varying properties of the subsurface or reservoirs. Some of the changes within a reservoir change the seismic velocity and thereby the seismic signals that travel through the reservoir. In characterizing these velocity changes, seismic signals are often influenced by changes in other properties of the reservoir such as fluid migration or the properties of the seismic sources of the signals. We investigate the impact of the perturbations in seismic source properties on time-lapse velocity estimation. We suggest a criterion for selecting seismic events that can be used in velocity analysis. This criterion depends on the dominant frequency of the signals, the centertime of the used time window in a signal, and the estimated relative velocity change. The criterion provides a consistent framework for monitoring changes in subsurface velocities using microseismic events and an ability to assess the accuracy of the velocity estimations.

Key words: time-lapse monitoring, coda wave interferometry, microseismicity

1 INTRODUCTION
Monitoring temporal changes within the Earth’s subsurface is a topic of interest in many areas of geophysics. These changes can result from an earthquake and its associated change in stress (Cheng et al., 2010), fluid injection or hydrofracturing (Davis et al., 2003), and oil and gas production (Zoback and Zinke, 2002). Some of the subsurface perturbations induced by these processes include temporal and spatial velocity changes, stress perturbations, changes in anisotropic properties of the subsurface, and fluid migration. Many of these changes span over a broad period of time and might even influence tectonic processes, such as induced seismicity (Zoback and Harjes, 1997). For example, Kilauea, Hawaii (which erupted in November 1975) is suggested to have triggered a magnitude 7.2 earthquake within a half hour of the eruption (Lipman et al., 1985). A seismic velocity perturbation of the subsurface leads to progressive time shifts across the recorded seismic signals. Various methods and data have been used to resolve the velocity perturbations. These methods include seismic coda wave interferometry (Snieder et al., 2002), doublet analysis of repeating microseismic and earthquake codas (Poupinet et al., 1984), time-lapse tomography (Vesnaver et al., 2003), and ambient seismic noise analysis (Cheng et al., 2010; Sens-Schönfelder and Wegler, 2006; Wegler and Sens-Schönfelder, 2007; Brenguier et al., 2008; Meier et al., 2010). Earthquake codas have higher sensitivity to the changes in the subsurface because multiple scattering allows these signals to sample the area of interest multiple times. However, there are inherent challenges in the use of these signals. Doublet analysis of the earthquake (microseismic) codas requires repeating events. Failure to satisfy the requirement that the events are identical can compromise the accuracy of the estimated velocity changes. In this study, we focus on the estimation of velocity changes using codas of repeating earthquakes that are not quite identical in their locations and source mechanisms.

Fluid-triggered microseismic events often are repeatable, but then events occur at slightly different positions with somewhat different source mechanisms (Sasaki and Kaieda, 2002; Godano et al., 2012). Imprints of the source perturbation and the velocity change on the seismic waveforms can be subtle, thereby retaining the similarity between seismic signals. Therefore, we will need to ask, how do the source location, source mechanism, and subsurface perturbations affect the estimated velocity changes? Snieder (2006)
shows that we can retrieve velocity changes from the coda of the waveforms recorded prior to and after the change. Robinson et al. (2011) develop a formulation using coda wave interferometry to estimate changes in source parameters of double-couple sources from correlation of the coda waves of doublets. Snieder and Vrijlandt (2005), using a similar formulation, relates the shift in the source location to the variance of the travel time perturbations between the doublet signals. In all these studies, the authors assume that the expected (average) change in travel time of the coda (due to either changes in the source locations or source mechanisms) is zero.

In this study, we investigate the impact of changes in source properties on the estimation of relative velocity changes. Knowledge of the impact of these perturbations on the estimated velocity change allows for a consistent framework for selecting pairs of earthquakes or microearthquakes used in velocity change analysis. This results in a more robust estimation of velocity change. In Section 2, we explore the theoretical relationships between the velocity changes and perturbations in the earthquake (microseismic) source properties. Following this section is a numerical validation of the theoretical results. We explain the implications and limitations of our results in Section 4. In the appendices, we explain the mathematical foundation of our results in this study.

2 MATHEMATICAL CONSIDERATION

In this section, we use the time-shifted cross-correlation (Snieder, 2002, 2006) to develop an expression for the average value of the time perturbation of scattered waves that are excited by sources with varying source properties. This perturbation is due to changes in the velocity of the subsurface and to changes in the source properties. These changes, we assume, may occur concurrently. Figure 1 is a schematic figure showing the general setup of the problem we are investigating. Two sources (S1 and S2) represent a microseismic doublet (repeating microseismic events). These events occur at different locations and can have different rupture patterns. We assume that these events can be described by a double couple. We investigate the ability of using the signals of these sources for time-lapse monitoring of velocity changes, assuming that these sources occur at different times. We express the signals of the two sources as unperturbed and perturbed signals, where the perturbation refers to any change in the signal due to changes within the subsurface and/or the source properties.

The unperturbed seismic signal $U(t)$ is given as

$$U(t) = A \sum_T U^{(T)}(t)$$

(1)

and the perturbed seismic signal $\hat{U}(t)$ is given as

$$\hat{U}(t) = \hat{A} \sum_T (1 + r^{(T)}) U^{(T)}(t - t^2_p),$$

(2)

where $A$ and $\hat{A}$ are the amplitudes of the unperturbed and perturbed source signals, respectively. These amplitudes represent the strengths of the sources. The recorded waves are a superposition of wave propagation along all travel paths as denoted by the summation over paths $T$. The change in the source focal mechanism only affects the amplitude of the wave traveling along each trajectory $T$ because the excitation of waves by a double couple is real (Aki and Richards, 2002). The change in the signal amplitudes - due to changes in the source mechanism angles - is defined by $r^{(T)}$ for path $T$, and $t^2_p$ is the time shift on the unperturbed signal due to the medium perturbation for path $T$. The change in the signal amplitudes along path $T$ depends on the source radiation angles. In this study, we assume that the medium perturbation results from the velocity change within the subsurface and changes in the source properties. The time-shifted cross-correlation of the two signals is given as

$$C(t_s) = \int_{t-t_w}^{t+t_w} U(t') \hat{U}(t' + t_s) \, dt',$$

(3)

where $t$ is the centertime of the employed time window and $2t_w$ is the window length. The normalized time-shifted cross-correlation $R(t_s)$ can be expressed as follows:

$$R(t_s) = \frac{\int_{t-t_w}^{t+t_w} U(t') \hat{U}(t' + t_s) \, dt'}{\left( \int_{t-t_w}^{t+t_w} U^2(t') \, dt' \int_{t-t_w}^{t+t_w+t_s} \hat{U}^2(t') \, dt' \right)^{1/2}}.$$

(4)

As shown in appendix A, the time-shifted cross-correlation has a maximum at a time lag equal to average time perturbation ($t_s = \langle t_p \rangle$) of all waves that
arrive in the used time window:
\[
\frac{1}{C(0)} \left. \frac{\partial C(t_s)}{\partial t_s} \right|_{(t_s - (t_p))} = 0. \tag{5}
\]
Equation (5) allows for the extraction of the average traveltime perturbation from the cross-correlation. In this study, the average of a quantity \( f \) is a normalized intensity weighted sum of the quantity \( \text{(Sniedder, 2006)}: \)
\[
\langle f \rangle = \frac{\sum_i A_i^2 f_i}{\sum_i A_i^2}.
\tag{6}
\]
where \( A_i^2 = \int U^T(t')^2 dt' \) is the intensity of the wave that has propagated along path \( T \).

We show in Appendices B and C that the expected value of the time perturbation and its variances are given by
\[
\langle t_p \rangle = -\left( \frac{\delta V}{V_0} \right) t
\tag{7}
\]
and
\[
\sigma_t^2 = \langle t_p^2 \rangle - \langle t_p \rangle^2 \approx \frac{D^2}{3V_0^2}.
\tag{8}
\]
In the above equations, \( \langle \delta V/V_0 \rangle \) is the average relative velocity change, \( D \) is the shift in the source location, and \( V_0 \) is the unperturbed velocity. Equation 7 suggests that the average time shift in the multiple scattered signals results only from the velocity changes within the subsurface. The variance of the time shifts depends, however, on the perturbations of the source location.

3 NUMERICAL VALIDATION

3.1 Description of numerical experiments

We test the equations in section 2 with the numerical simulation using Foldy’s multiply scattering theory (Foldy, 1945) described by (Groenenboom and Snieder, 1995). The theory models multiple scattering of waves by isotropic point scatterers. We conduct our numerical experiments using a circular 2D geometry (Figure 2) with point scatterers surrounded by 96 receiver stations. We uniformly assign the imaginary component of the scattering amplitude \( |Im A| = 4 \) to all the scatterers. In 2D, this is the maximum scattering strength that is consistent with the optical theorem that allows for conservation of energy (Groenenboom and Snieder, 1995).

We assume an acoustic propagating wavefield which is a sum of the direct wavefield \( U_0(r) \) and the scattered wavefield from all the scatterers:
\[
U(r) = U_0(r) + \sum_{i=1}^{n} G^{(0)}(r, r_i) A_i U(r_i),
\tag{9}
\]
where the first term is the direct wave and the scattered wavefield is described by the second term. The Green’s function at \( r \) due to scatterer \( i \) at position \( r_i \) is \( G^{(0)}(r, r_i) \) while \( U(r_i) \) is the incoming wavefield at scatterer \( i \). The scattering amplitude of each scatterer is given by \( A_i \). The direct wave is modulated by the far-field P-wave radiation pattern \( F^P \):
\[
U_0(r) = F^P G^{(0)}(r, r_s),
\tag{10}
\]
with \( r_s \) the source location. In 2D, where the take-off direction is restricted within the 2D plane, \( F^P \) is given as (Aki and Richards, 2002)
\[
F^P = \cos \lambda \sin \delta \sin 2(\psi - \phi) - \sin \lambda \sin 2\delta \sin^2 (\psi - \phi),
\tag{11}
\]
where \( \psi \) is the azimuth of the outgoing wave and \( \lambda, \delta, \) and \( \phi \) are the source parameters (rake, dip and strike, respectively). Sources are located at the center of the scattering area. The source spectrum has a dominant frequency \( f_d \) of approximately 30 Hz and a frequency range of 10-50 Hz. The source spectrum tapers off at the frequency extremes by a cosine taper with a length given by half of the bandwidth. We assume a reference velocity \( V_0 = 3500 \text{ m/s} \). Because the model we are using is an isotropic multiple scattering model, the transport mean free path is the same as the scattering mean free path: \( l^* = l \). The scattering mean free path \( l^* \) is given by (Busch et al., 1994; Groenenboom and Snieder, 1995)
\[
l^* = \frac{k_o}{\rho |Im A|}.
\tag{12}
\]
where \( k_o \) is the wavenumber of the scattered signal and \( \rho \) is the scatterer density. For our model, the mean free path \( l^* \) is approximately 12.2 km. There is no intrinsic attenuation in the numerical model.

We generate multiple scattered signals, which are recorded at the receivers, using the numerical model in Figure 2. These signals are generated with a reference model defined by the following reference parameter values: the source radiation parameters \( \phi = 0^\circ, \lambda = 0^\circ, \delta = 90^\circ \); change in medium velocity \( \Delta V = 0 \text{ m/s} \); and shift in the source location \( D = 0 \text{ m} \). We refer to signals generated by this reference model as the reference signals. In order to understand the effect of the perturbation of these parameters on velocity change estimation, we also generate synthetic signals from the perturbed version of the model. The perturbed model consists of perturbation of either the source locations, source radiation parameters, the medium velocity, or a combination of these. Synthetic signals from the reference and the 0.4% velocity perturbed models are shown in Figure 3 with zoom insets showing the stretching of the wavefield by the velocity perturbation. The result of the velocity perturbation on the signals is a progressive time shift of the arriving seismic phases in the signals.

3.2 Data processing

To estimate the velocity perturbations or possible velocity change imprints on the synthetic signals due to the perturbation of the source location or its radiation
properties, we use the stretching algorithm of (Hadziioannou et al., 2009) who demonstrate the stability and robustness of the algorithm relative to the moving time-window cross-correlation of (Snieder et al., 2002) and the moving time-window cross-spectral analysis of (Poupinet et al., 1984). Both algorithms satisfy the relative velocity change equation (Snieder, 2006):

$$\left\langle \frac{t_p}{t} \right\rangle = -\epsilon,$$  

where $\epsilon = (\delta V/V_o)$ is the relative velocity change.

In the stretching algorithm, we multiply the time of the perturbed signal with a stretching factor $(1 - \epsilon)$ and interpolate the perturbed signal at this stretched time. The time window we use in all our analysis is given by the black bold line in Figure 3. We then stretch the perturbed signal at a regular interval of $\epsilon$ values. The range of the $\epsilon$ values can be arbitrarily defined or predicted by prior information on the range of changes in the subsurface velocity. To resolve the value of $\epsilon$, we use an $L_2$ objective function rather than the cross-correlation algorithm as suggested by (Hadziioannou et al., 2009). For events of equal magnitude ($A = \hat{A}$), the objective function is

$$R(\epsilon) = ||\hat{U}(t(1 - \epsilon)) - U(t)||_2,$$  

where $||...||_2$ is the $L_2$ norm. Figure 4 shows the objective function for the case of a 0.4% velocity change. The minimum of the objective function depends on the amplitude changes between the two signals and on the travel time perturbations due to velocity changes and shifts in the source location. The signals have uniform magnitudes. The amplitude changes between the signals are due to changes in the orientation of the source angles.

The error in the estimated relative velocity change $c_{\delta v}$ is given by

$$c_{\delta v} \leq \frac{\sigma_U}{2\pi f_d A t},$$  

where $f_d$ is the dominant frequency, $t$ is the centertime of the signal, $A$ is the amplitude of the signals, and $\sigma_U$ is the standard deviation of the recorded waveforms. The derivation of the error equation (equation 15) is given in Appendix D. The error associated with the velocity change depends on additive noise in the signals and on differences in the signals both in amplitude and in phase due to perturbation in source properties.

### 3.3 Effect of perturbation of source properties on the estimated velocity change

To understand the effect of the changes in the source properties on the estimation of the relative velocity changes, we conduct our numerical experiment over a range of parameter changes. We perturb the source location and the orientation of the source angles. The perturbation of the source radiation parameters is characterized by the weighted root mean square change in source parameters $\langle r \rangle$ (Robinson et al., 2007):

$$\langle r \rangle = -\frac{1}{2}(4\Delta \phi^2 + \Delta \lambda^2 + \Delta \delta^2),$$  

where...
Estimating velocity changes with dissimilar sources

Figure 4. The objective function $R(\epsilon)$ as a function of the stretch factor $\epsilon$. The objective function is minimum for $\epsilon = 0.4\%$, which corresponds to the time velocity change.

Figure 5. Estimated relative velocity change for model velocity change of $\langle \delta V/V_0 \rangle = 0.1\%$ (blue), $\langle \delta V/V_0 \rangle = 0.2\%$ (red), $\langle \delta V/V_0 \rangle = 0.3\%$ (black), and $\langle \delta V/V_0 \rangle = 0.4\%$ (green).

where $\Delta \phi$ is the change in strike, $\Delta \lambda$ is the change in rake, and $\Delta \delta$ is the change in dip. These changes represent the angle differences between the two sources (doublet).

Figures 5 and 6 show the estimated velocity changes due to the perturbation of medium velocity and the source properties (location and radiation parameters), respectively. For Figure 5, we generate signals with the following perturbation in the model velocity $\langle \delta V/V_0 \rangle$: 0.1%, 0.2%, 0.3%, and 0.4%. In these models, we keep the source parameters unchanged. Using the stretching method, we are able to recover the velocity changes we impose in the model from the codas of each of the perturbed signals and that of the reference signal (Figure 5). The method accurately estimates the model velocity change in all the receivers. We also generate signals with only perturbations in the source locations and mechanisms. In this case, the true velocity change is zero. Figure 6 shows that the estimated relative velocity changes $\langle \delta V/V_0 \rangle$ are near the true value $\langle \delta V \rangle = 0$ for models with perturbations of either the source location or the source radiation parameters. The velocity change estimated from individual stations varies around zero, but with a shift in the source location of $D = 0.143\lambda_d$, with $\lambda_d$ the dominant wavelength and source angle perturbations as large as $\Delta \phi = 20^\circ$, $\Delta \lambda = 20^\circ$, and $\Delta \delta = 20^\circ$ ($r = -0.366$), the magnitude of these variations is smaller than $1/20$th of the typical velocity changes inferred from seismic signals (Figure 6). These variations in the velocity change inferred from different stations can be used to estimate the errors in the estimated velocity change. These results agree with equation (7) which predicts that the average value of time shifts in the perturbed signal results only from changes in the medium velocity and is not affected by changes in source properties. We will need to know how effectively we can resolve the velocity changes in our model in the presence of the other model parameter perturbations.

Figure 6. Estimated relative velocity change due to perturbation in the source location and the source radiation. A. The estimated velocity change for perturbed source location (divided by the dominant wavelength; inset in the top right) and B. The estimated velocity change caused by changes in source radiation angles (inset in the top right). The value of $r$ is given by equation 14.
3.4 Limiting regimes of the estimations

To investigate the extent of the perturbation in the source location and source radiation perturbations that can be allowed in the estimation of relative velocity changes, we generate synthetic signals with models having a 0.1% relative velocity change and various perturbations of the source parameters. The values of the source parameter perturbations are given in Tables 1 and 2. Figure 7 shows the estimated relative velocity changes from signals generated from sources at different locations. The figure shows that we can recover the model relative velocity change of 0.1% using doublets (two sources) within a sphere of radius \( \lambda_d/4 \), where \( \lambda_d \) is the dominant wavelength of the seismic signal which is approximately 140 m. This is consistent with the criterion we derived in Appendix E, which predicts that for an accurate estimation of the subsurface velocity change, the shift in the source location has to satisfy

\[
\frac{D}{\lambda_d} < \sqrt{2 \left( \frac{\delta V}{V_0} \right) f_d t}, \quad (17)
\]

where \( f_d \) is the dominant frequency of the signals, \( \langle \delta V/V_0 \rangle \) is the average velocity change in the subsurface, and \( t \) is the centertime of the processed signal. The criterion is derived from a comparison of the phase changes due to velocity changes with those due to shifts in the source location. For the results in Figure 7, our model parameters are \( \langle \delta V/V_0 \rangle = 0.001 \), \( t = 10s \), and \( f_d \approx 25\text{Hz} \). With these values, the constraint on the source location shift for Figure 7 is

\[
\frac{D}{\lambda_d} < 0.35. \quad (18)
\]

Figure 7 shows that for \( D/\lambda_d \geq 0.3 \), the estimated velocity change deviates significantly from the real velocity change; this is in agreement with equation (18). The criterion in equation (17) imposes a constraint on the spacing requirements for the source locations of the doublets used for time-lapse velocity change monitoring with microearthquakes. Alternatively, equation (17) gives the magnitude of a velocity change that is resolvable with a given shift in the source location. According to equation (17), the allowable source separation increases with the centertime \( t \) of the employed time window. This is due to the fact that the imprint of the velocity change is more pronounced as the waves have propagated over a greater distance through the perturbed medium. However, signals at later times in the coda are more affected by the presence of additive noise because the signal-to-noise ratio usually decreases towards the late coda.

We also investigate the effect of the source radiation properties on the estimated velocity change of the medium of interest. Figure 8 shows the estimated velocity changes from a model with 0.1% velocity change using sources with perturbed radiation angles (measured by \( \langle r \rangle \)). The values of the perturbed source radiation angles are given in Table 2. In Figure 8A, the estimated

---

**Table 1.** Modeling parameters for shift in the source location.

<table>
<thead>
<tr>
<th>Case</th>
<th>( \frac{\Delta V}{V_0} )</th>
<th>( D (\lambda_d) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Case 1</td>
<td>0.1</td>
<td>0.0274</td>
</tr>
<tr>
<td>Case 2</td>
<td>0.1</td>
<td>0.0547</td>
</tr>
<tr>
<td>Case 3</td>
<td>0.1</td>
<td>0.0820</td>
</tr>
<tr>
<td>Case 4</td>
<td>0.1</td>
<td>0.1094</td>
</tr>
<tr>
<td>Case 5</td>
<td>0.1</td>
<td>0.1367</td>
</tr>
<tr>
<td>Case 6</td>
<td>0.1</td>
<td>0.1642</td>
</tr>
<tr>
<td>Case 7</td>
<td>0.1</td>
<td>0.1918</td>
</tr>
<tr>
<td>Case 8</td>
<td>0.1</td>
<td>0.2197</td>
</tr>
<tr>
<td>Case 9</td>
<td>0.1</td>
<td>0.2477</td>
</tr>
<tr>
<td>Case 10</td>
<td>0.1</td>
<td>0.2760</td>
</tr>
<tr>
<td>Case 11</td>
<td>0.1</td>
<td>0.3047</td>
</tr>
<tr>
<td>Case 12</td>
<td>0.1</td>
<td>0.3337</td>
</tr>
<tr>
<td>Case 13</td>
<td>0.1</td>
<td>0.3627</td>
</tr>
<tr>
<td>Case 14</td>
<td>0.1</td>
<td>0.3916</td>
</tr>
<tr>
<td>Case 15</td>
<td>0.1</td>
<td>0.4202</td>
</tr>
</tbody>
</table>
velocity change at the individual stations progressively deviates from the true velocity change of 0.1% with increasing change in the orientations of the source angles. This deviation is due to the decorrelation between the perturbed and the unperturbed signals as shown in Figure 8B, which shows the maximum normalized cross-correlation of the codas within the processed time window. With an increasing change in the orientation of the sources, the maximum cross-correlation value of the waves excited by the doublets decreases. However, for source angle perturbations as large as \( \Delta \phi = 28^\circ \), \( \Delta \lambda = 28^\circ \), and \( \Delta \delta = 28^\circ \) corresponding to \( |\langle r \rangle| = 0.72 \) (Figure 8A), the maximum deviation from the 0.1% model velocity change is approximately 0.01%. This is a small change compared to velocity changes resolved from seismic signals in practice. Preliminary results on the velocity change can be used to pick events satisfying equation (17) for an accurate estimation of the velocity changes. Using doublets that do not satisfy the constraint result in an inaccurate estimate of the velocity change.

A significant change in the source mechanism of double couple sources can introduce a bias in the estimation of relative velocity change. This bias is due to the decorrelation of the perturbed and unperturbed signals which lowers the accuracy of the estimated velocity change. As shown in Figures 6B and 8A, some of the stations underestimate while others overestimate the velocity change. However, this bias is negligible for the typical velocity changes resolved from seismic signals in practice. This result permits the use of sources

4 DISCUSSION AND CONCLUSION

In this study, we investigate the influence of perturbation in source properties (location and radiation) on the estimation of velocity changes. These velocity changes are extracted from multiply scattered signals (codas) of repeating events. We show that we can resolve accurate values of relative velocity changes if the shift in the source location satisfies the constraint (equation 17). This constraint depends on the dominant frequency of the signal, the estimated relative velocity change, and the centertime of the employed time window. This places a restriction on the relative event locations that can be used to estimate the relative velocity change of the subsurface. However, to use this constraint, we need to know the magnitude of the relative velocity change we seek to measure. Preliminary results on the velocity change can be used to pick events satisfying equation (17) for an accurate estimation of the velocity changes. Using doublets that do not satisfy the constraint result to an inaccurate estimate of the velocity change.

Table 2. Modeling parameters for source radiation perturbation.

<table>
<thead>
<tr>
<th>Case</th>
<th>( \Delta V ) (%)</th>
<th>( \Delta \phi ) (\°)</th>
<th>( \Delta \lambda ) (\°)</th>
<th>( \Delta \delta ) (\°)</th>
<th>(-\langle r \rangle)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Case 1</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Case 2</td>
<td>0.1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0.0037</td>
</tr>
<tr>
<td>Case 3</td>
<td>0.1</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>0.0146</td>
</tr>
<tr>
<td>Case 4</td>
<td>0.1</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>0.0329</td>
</tr>
<tr>
<td>Case 5</td>
<td>0.1</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>0.0585</td>
</tr>
<tr>
<td>Case 6</td>
<td>0.1</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>0.0914</td>
</tr>
<tr>
<td>Case 7</td>
<td>0.1</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>0.1316</td>
</tr>
<tr>
<td>Case 8</td>
<td>0.1</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>0.1791</td>
</tr>
<tr>
<td>Case 9</td>
<td>0.1</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>0.2339</td>
</tr>
<tr>
<td>Case 10</td>
<td>0.1</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>0.2961</td>
</tr>
<tr>
<td>Case 11</td>
<td>0.1</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>0.3655</td>
</tr>
<tr>
<td>Case 12</td>
<td>0.1</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>0.4423</td>
</tr>
<tr>
<td>Case 13</td>
<td>0.1</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>0.5264</td>
</tr>
<tr>
<td>Case 14</td>
<td>0.1</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>0.6178</td>
</tr>
<tr>
<td>Case 15</td>
<td>0.1</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>0.7165</td>
</tr>
</tbody>
</table>

Figure 8. Effect of source angle perturbation on estimated velocity change. A. The estimated relative velocity changes are from a 0.1% model velocity change and various source radiation perturbations (Table 2). B. The decorrelation of the doublets, due to source angle perturbations, measured by maximum cross-correlation values. Stations SW, SE, NE, and NW positions are given in Figure 2.
of different orientations for the estimation of velocity changes, provided that the maximum cross-correlation of the source signals is greater than 0.7 as shown in Figure 8B. For a consistent estimate of the velocity change, using multiple stations is useful to ascertain the accuracy of the estimated velocity change in an isotropic subsurface.

The theory presented in this study is based on a number of simplifications and assumptions. First, we assume a uniform velocity change across our model. For the case of a localized isotropic velocity change, the resolved velocity change is a fraction of the local velocity change, where the fraction is dependent on the amount of time the codas spend within the perturbed region relative to the unperturbed region. Also, we assume that the velocity is isotropic. In an anisotropic velocity medium, we need to include the dependence of the velocity on the direction of propagation.

In this study, we assume that the scatterer density is uniform in all directions from the source. We also ignore changes in the scattering properties which might include shifts in scatterer locations (Niu et al., 2003) and changes in the scattering strength of the scatterers. These changes in the scattering properties can be due to changes in fluid properties such as fluid migration or opening and closing of fractures and pre-existing faults.

If the shifts in the scatterer location are random, the average travel time perturbation due to scatterer location shift is zero. However, if the shifts in the scatterer locations are non-random or directional, the average time perturbation due to scatterer location shift over all take-off angles is a non-zero mean traveltime change. The scattered signals lag behind while traveling through a higher scatterer density region compared to a lower scatterer density region. These introduce a bias in the estimated relative velocity changes if the changes in scatterer properties or density are significant.

We used point scatterers in our numerical modeling even though in the real world, scattering can be caused by faults, fractures, horizontal or dipping layers. The employed modeling uses scalar waves, hence it does not account for mode conversions of elastic waves (for example, P-to-S or S-to-P and surface waves) that might result due to the presence of layers, free surface, and fractures. The coda is usually dominated by S wave (Aki and Chouet, 1975; Snieder, 2002), hence the mode conversions between P and S waves do not dominate the details of the scattering processes.

**APPENDIX A: EXTRACTING TIME SHIFT FROM THE CROSS-CORRELATION**

This appendix shows the extraction of the average time perturbation from the time-shifted cross-correlation. Using the time-shifted cross-correlation, we show that the maximum value of the cross-correlation occurs at a time lag which is equal to the average time perturbation resulting from changes either in the subsurface velocity, source locations, or source focal mechanisms. Re-expressing time-shifted cross correlation function (equation 3) gives

$$C(t_s) = \sum_{T} (1 + r^{(T)}) \int_{t-t_w}^{t+t_w} (U^T(t')U^T(t' + t_s - t_p^T)) \, dt'.$$

(A1)

The time-shifted cross-correlation has double sum contributions from the diagonal terms $\sum_{T=T'}$ and the crossterms $\sum_{T \neq T'}$ due to the product of $U(t)$ and $U(t')$ which individually involve summation over all trajectories. The summation in equation A1, only accounts for the contributions from the diagonal terms ($\sum_{T,T'} = \sum_{T=T'} + \sum_{T \neq T'} \rightarrow \sum_{T=T'}$), because we ignore the crossterms $\sum_{T \neq T'}$. The crossterms depend on the bandwidth of the signals $\Delta(f)$ and the length of the cross-correlation window $2t_w$ (Snieder, 2006):

$$\frac{|\text{crossterms}|}{|\text{diagonal terms}|} = \sqrt{\frac{1}{\Delta f 2t_w}}. \quad (A2)$$

The crossterms become negligible with an increasing length of time window used for the time window.

Using a Taylor’s series expansion of $U^T(t' + t_s - t_p^T)$ in $t_s - t_p^T$,

$$C(t_s) = \sum_{T} (1 + r^{(T)}) \int_{t-t_w}^{t+t_w} \left( U^T(t')^2 + \frac{(t_s - t_p^T)}{2} U^T(t')U^T(t') + \frac{(t_s - t_p^T)^2}{2} U^T(t')U^T(t') \right) \, dt'.$$

(A3)

The contribution of the first time derivative $\dot{U}^T(t')$ term can be reduced on the end-points contributions of the signals. Since we taper the windowed coda to eliminate edge effects these end-point contributions vanish:

$$\int_{t-t_w}^{t+t_w} (t_s - t_p^T) U^T(t') \dot{U}^T(t') \, dt' = \frac{1}{2} U^{T2}(t) \bigg|_{t-t_w}^{t+t_w} = 0,$$

(A4)

because the tapered signals at the end-points are zero (Snieder, 2006). Using integration by parts, equation A3 can be re-expressed as

$$C(t_s) = \sum_{T} (1 + r^{(T)}) \times \int_{t-t_w}^{t+t_w} (U^T(t')^2 - \frac{(t_s - t_p^T)^2}{2} U^T(t')^2) \, dt'.$$

(A5)

The time-shifted cross correlation is maximum when it’s first derivative with respect to the lagtime $t_s$ is zero. The first derivative of the cross-correlation function with re-
spect to $t_s$ is given as
\[
\frac{\partial C(t_s)}{\partial t_s} = -\sum_T (1 + r^{(T)})(t_s - t_p^T) \int_{t-t_w}^{t+t_w} (U^T(t')^2) \, dt'.
\]

Using $C(0) = \sum_T \int_{t-t_w}^{t+t_w} U^T(t')^2 \, dt'$, the normalized derivative of the time-shifted cross-correlation is given by
\[
\frac{1}{C(0)} \frac{\partial C(t_s)}{\partial t_s} = -\sum_T (1 + r^{(T)})(t_s - t_p^T) \int_{t-t_w}^{t+t_w} (U^T(t')^2) \, dt'.
\]

We next define the mean frequency $\bar{\omega}^2$ as
\[
\bar{\omega}^2 = \frac{\sum_T \int_{t-t_w}^{t+t_w} (U^T(t')^2) \, dt'}{\sum_T \int_{t-t_w}^{t+t_w} (U^T(t'))^2 \, dt'}.
\]

Therefore, using equation (6)
\[
\frac{1}{C(0)} \frac{\partial C(t_s)}{\partial t_s} = -(1 + \langle r \rangle)(t_s - \langle t_p \rangle) \bar{\omega}^2.
\]

We assume that the amplitude perturbations due to the change in the source mechanism and the traveltime perturbations are independent. In that case, using equation 6, equation A9 becomes
\[
\frac{1}{C(0)} \frac{\partial C(t_s)}{\partial t_s} = -(1 + \langle r \rangle)(t_s - t_p^T) \bar{\omega}^2.
\]

The right hand side vanishes when
\[
t_s - \langle t_p \rangle = 0.
\]

This shows that the time shift $t_s$ where the cross-correlation is maximum is the average of the travel-time perturbations.

**APPENDIX B: THE TIME PERTURBATION DUE TO A PERTURBED SOURCE**

From Figure 1, the traveltime $t_T$ for the signal along path $T$ from the unperturbed source to the first scatterer along path $T$ is given by
\[
t_T = \frac{L_T}{V_o}.
\]

where $V_o$ is the unperturbed medium velocity. Also the traveltime $t_{T'}$ for the signal along path $T'$ from the perturbed source to the same first scatterer is given by
\[
t_{T'} = \frac{L_{T'}}{V_o}.
\]

We assume that the signals from both sources after scattering by the first scatterer travel along the same path (Figure 1). We define $L_T = L_{T'} + \delta L$ and $V = V_0 + \delta V$, where $\delta L = -\langle \mathbf{r}_T \cdot \mathbf{D} \rangle$, with $\mathbf{D}$ the perturbation of the source location and $\mathbf{r}_T$ the take-off direction from the first source to the scatterer. The takeoff direction from the source $\mathbf{r}_T$ in spherical coordinates is
\[
\mathbf{r}_T = \left(\frac{\cos \psi \sin \theta}{\cos \theta}, \frac{\sin \psi \sin \theta}{\cos \theta}, \cos \theta\right),
\]

where $\psi$ and $\theta$ are colatitude and longitude in spherical coordinates. The traveltime for the signal along path $T$ from the second source to the first scatterer can be re-expressed as
\[
t_{T'} = \frac{L_{T'}}{V_0 + \delta V}
\]

\[
\simeq \left(\frac{1}{V_o} - \frac{\delta V}{V_o^2} + \ldots \right) L_T - \left(\frac{1}{V_o} - \frac{\delta V}{V_o^2} + \ldots \right) \cdot \left(\mathbf{r}_T \cdot \mathbf{D}\right).
\]

Ignoring the terms of second order or higher, equation B5 gives,
\[
t_{T'} = \frac{L_T}{V_o} - \frac{\delta V}{V_o} \left(\frac{L_T}{V_o} - \frac{\langle \mathbf{r}_T \cdot \mathbf{D} \rangle}{V_o}\right),
\]

\[
= t_T + t_{pv} + t_{pl}.
\]

Therefore, the time perturbation along path $T'$ is given as
\[
t_{T'} = t_{T'} - t_T = t_{pv} + t_{pl},
\]

where $t_{pv} = $ time shift due to velocity change and $t_{pl} = $ time shift due to shift in source .

We need to derive the expression for $\langle t_p \rangle$. With equation 6 and equation B8, the average time perturbation is given

\[
\langle t_p \rangle = \frac{\sum_T A_p^2 \int_{t-t_w}^{t+t_w} \frac{1}{V_o} \, dt}{\sum_T A_p^2} = \frac{-\sum_T A_p^2 \left(\left(\frac{\delta V}{V_o^2}\right) (\bar{\omega}^2) + \frac{\mathbf{r}_T \mathbf{D}}{V_o}\right)}{\sum_T A_p^2},
\]

\[
= \left\langle \delta V \right\rangle \frac{\int_T \frac{L_T}{V_o} \, d\Omega}{\int_T \frac{L_T}{V_o} \, d\Omega} + \frac{\int_T (\mathbf{r}_T \mathbf{D})/V_o \, d\Omega}{\int_T \frac{L_T}{V_o} \, d\Omega},
\]

where $\int_T \ldots \, d\Omega$ denotes an integration over all take-off angles. In 3D, the integration limits of $d\theta$ and $d\psi$ are $[0, \pi]$ and $[0, 2\pi]$, respectively. Since $\int_T \frac{L_T}{V_o} \, d\Omega = 0$ and $\int_T \frac{L_T}{V_o} \, d\Omega/\int_T \frac{L_T}{V_o} \, d\Omega = t$, where $t$ is the traveltime of the scattered ray from the source to the receiver along path $T$, equation B9 reduces to

\[
\langle t_p \rangle = -\left\langle \frac{\delta V}{V_o} \right\rangle t.
\]

Hence to first order in $D$, the average traveltime perturbation depends only on the velocity changes within the explored medium.
APPENDIX C: VARIANCE OF TIME PERTURBATION

The variance of the travel time perturbation, using equation 6, is given by

\[ \sigma_t^2 = \frac{\sum_T A_T^2 (t_p - \langle t_p \rangle)^2}{\sum_T A_T^2} = (\langle t_p^2 \rangle - \langle t_p \rangle^2). \]  

(C1)

where using expression B-8 \( \langle t_p^2 \rangle \) is given by

\[ \langle t_p^2 \rangle = \frac{\sum_T A_T^2 \left( -\frac{\delta V}{V_o} \left( \frac{L_T}{V_o} \right) t_p - \frac{\dot{r}_T \cdot D}{V_o} \right)^2 + 2 \left( \frac{\delta V}{V_o} \right) \left( \frac{L_T}{V_o} \right) \left( \frac{\dot{r}_T \cdot D}{V_o} \right)}{\sum_T A_T^2}. \]  

(C2)

Expanding equation (C2) gives

\[ \langle t_p^2 \rangle = \sum_T A_T^2 \left( \frac{\delta V}{V_o} \right)^2 \left( \frac{L_T}{V_o} \right)^2 + \left( \frac{\dot{r}_T \cdot D}{V_o} \right)^2 \]

\[ + 2 \left( \frac{\delta V}{V_o} \right) \left( \frac{L_T}{V_o} \right) \left( \frac{\dot{r}_T \cdot D}{V_o} \right) \]  

(C3)

In equation (C3),

\[ \frac{\int_P (\dot{r}_T \cdot D) \, \frac{d^2 \Omega}{d\Omega}}{\int_P \frac{d^2 \Omega}{d\Omega}} = \frac{D^2}{C}, \]  

(C4)

where \( C = 1, 2, \) or 3 which specifies the dimension of the problem. Therefore, in 3D,

\[ \langle t_p^2 \rangle = \left( \frac{\delta V}{V_o} \right)^2 \left( \frac{L_T}{V_o} \right)^2 + \frac{2 \left( \frac{\delta V}{V_o} \right) \left( \frac{L_T}{V_o} \right) \left( \frac{\dot{r}_T \cdot D}{V_o} \right)}{3V_o^2}. \]  

(C5)

Combining equations (C5) and (B10), the total variance of the time perturbation is

\[ \sigma_t^2 = \frac{D^2}{3V_o^2}. \]  

(C6)

In the absence of additive noise, the variance of the traveltime perturbation thus depends only on the shift in the source location. With the estimate of the subsurface velocity, we can estimate the shift in the source location from equation C6.

APPENDIX D: ERROR ESTIMATION

We estimate the error associated with the estimated relative velocity change using the data residuals from the \( L_2 \) norm. Using a Taylor series expansion of \( \dot{U}(t + t_p) \) in \( t_p \),

\[ \dot{U}(t + t_p) \simeq \dot{U}(t)(1 - i\omega_d t_p) \]  

(D1)

where \( \omega_d \) is the \( 2\pi f_d \) with \( f_d \) the dominant frequency of the signal. Including additive errors \( \delta U(t) \) in the data with standard deviation \( \sigma_U \), equation D1 gives

\[ \dot{U}(t + t_p) + \delta U(t) \simeq \dot{U}(t)(1 - i2\pi f_d(t_p + \delta t)). \]  

(D2)

where \( \delta t \) is the error of the traveltme perturbation due to the error in the data \( \dot{U}(t + t_p) \). This error leads to an error in the estimated lagtime from the cross-correlation of the perturbed and unperturbed signals. Therefore, the relationship between the data error and the error in the time perturbation is

\[ \sigma_u \simeq -iU(t)2\pi f_d \sigma_t. \]  

(D3)

where \( \sigma_t \) is the standard deviation of the error in the traveltme perturbation. Therefore, the error in the time perturbation between the perturbed and unperturbed signals is

\[ \sigma_t = ||\dot{t}_p - t_p|| \leq \frac{\sigma_u}{|||U|||2}. \]  

(D4)

where \( \dot{t}_p \) and \( t_p \) are the estimated and the exact time perturbations, respectively, and the M is \(-iU(t)2\pi f_d \).

Combining equations 7 and D4, the error in the estimated relative velocity change \( \varepsilon_{\delta v} \) is the time-normalized time perturbation error

\[ \varepsilon_{\delta v} = \frac{\sigma_t}{t} \leq \frac{\sigma_u}{|||U|||22\pi f_d t}. \]  

(D5)

In practice, \( t \) is the center time of the used time window, \( |||U|||2 \) is the amplitude of the data.

The error equation (equation D5) depends on the dominant frequency of the signal, the length of the signals, and the amplitude difference between the signals \( U(t) \) and \( \dot{U}(t) \) which is normalized by the amplitude of \( U(t) \). The error in the data \( \sigma_u \) is due to any dissimilarity between the two signals \( \dot{U}(t) \) and \( U(t) \) resulting from either shift in the source location or the presence of additive noise.

APPENDIX E: COMPARATIVE TIME SHIFT BETWEEN CHANGES IN VELOCITY AND SOURCE LOCATION

In this section, we compare phase shifts due to the shift in the source location to the phase shifts resulting from velocity change within the subsurface. If the phase of the wave that travels over a distance \( r \) from a source to a scatterer is \( \exp(-ik \cdot r_T) \), the phase change due to shift in the source location along path \( T \) is

\[ \exp(-ik \cdot r_T) = \exp(-ikD \cos \theta_T), \]  

(E1)

where \( \theta_T \) is the angle between the take-off ray of path \( T \) and the shift in the source location \( D \), and \( k = 2\pi/\lambda \). For \( D/\lambda < 1 \), we can approximate equation E-1 as

\[ \exp(-ikD \cos \theta_T) \simeq 1 - ikD \cos \theta_T - \frac{1}{2}(kD \cos \theta_T)^2. \]  

(E2)

The average value of the phase changes due to the shift in the source location is

\[ \langle \exp(-ikD \cos \theta_T) \rangle \simeq 1 - ikD(\cos \theta_T) - \frac{1}{2}(kD)^2(\cos^2 \theta_T), \]  

(E3)
where \( \langle \cos \theta_T \rangle = 0 \), assuming we sum over all angles. For equal contribution from all take-off angles in 2D (the numerical simulations in section 3 are in 2D), \( \langle \cos^2 \theta_T \rangle = \frac{1}{2} \). Therefore,

\[
\langle \exp(-ikD \cos \theta_T) \rangle \approx 1 - \frac{1}{4}k^2D^2. \tag{E4}
\]

Also, if the phase of the wave that travels over a time \( t \) is \( \exp(-i\omega t) \), the phase change due to the change in the medium velocity is

\[
\exp(-i\omega \Delta t) \approx 1 - i\omega \Delta t - \frac{1}{2}(\omega \Delta t)^2, \tag{E5}
\]

where \( \Delta t \) is the time shift due to velocity change. The second order terms contribute to the variance of the phase change. Therefore, for an accurate estimation of the velocity change,

\[
\frac{1}{4}k^2D^2 < \frac{1}{2}\omega^2\Delta t^2. \tag{E6}
\]

Equation E-6 implies that

\[
\frac{D}{\lambda} < \sqrt{2f|\Delta t|}. \tag{E7}
\]

But the average value of time shift due to velocity change \( \langle \Delta t \rangle \) is

\[
\langle \Delta t \rangle = -\left\langle \frac{\delta V}{V_0} \right\rangle t. \tag{E8}
\]

Therefore, equation (E-7) reduces to

\[
\frac{D}{\lambda} < \sqrt{2f}\left\langle \frac{\delta V}{V_0} \right\rangle t. \tag{E9}
\]

Equation E9 shows that for an accurate estimation of relative velocity changes, the shift in the source location \( \Delta t \) has to satisfy equation E9. For practical purposes, \( \lambda \) and \( f \) can be defined as the dominant wavelength and frequency of the processed signal, respectively. Also, \( t \) can be assigned as the centertime of the used time window.

ACKNOWLEDGMENTS

We are grateful for the financial support of the Department of Energy (DOE) through grant DE-EE0002758.

REFERENCES


Robinson, D. J., R. Snieder, and M. Sambridge, 2007, Using coda wave interferometry for estimating the


Feasibility of inverting compaction-induced traveltime shifts for reservoir pressure

Steven Smith(1) and Ilya Tsvankin(2)
Colorado School of Mines, Center for Wave Phenomena, Golden Colorado

(1) email: smithsh@dix.mines.edu
(2) email: ilya@dix.mines.edu

ABSTRACT

Pore-pressure variations inside producing reservoirs result in excess stress and strain that cause time shifts of reflected waves. Inversion of seismic data for pressure changes requires better understanding of the dependence of compaction-induced time shifts on reservoir pressure reduction. Here, we investigate pressure-dependent behavior of P-, S-, and PS-wave time shifts from reflectors located above and below a rectangular reservoir embedded in a homogeneous half-space. Our geomechanical modeling algorithm generates the excess stress/strain field and the stress-induced stiffness tensor as linear functions of reservoir pressure. Analysis of time shifts obtained from full-waveform synthetic data shows that they vary linearly with pressure for reflectors above the reservoir, but become nonlinear for reflections from the reservoir or deeper interfaces. Time-shift misfit curves computed with respect to noise-contaminated data from a reference reservoir for a wide range of pressure reductions display well-defined minima. We also evaluate the influence of the reservoir width and third-order stiffness coefficients on time shifts because these parameters will be fixed in the inversion for reservoir pressure. Our results indicate that stable pressure estimation requires including multicomponent time shifts from reflectors below the reservoir and applying a nonlinear global inversion algorithm.

Key words: geomechanics, seismic modeling, inversion, stress-induced anisotropy, converted waves, shear waves, time-lapse, compacting reservoir, transverse isotropy, VTI

Introduction

Compaction-induced seismic travelt ime shifts can potentially be inverted for pressure and fluid distributions inside a producing reservoir. Such an inversion contributes to the understanding of how fluids are moving (sweeping) through a reservoir, of levels of intercompartment pressure communication, and whether fluid is produced away from wells (Greaves and Fulp, 1987; Landro, 2001; Lumley, 2001; Calvert, 2005; Hodgson et al., 2007; Wikel, 2008). Knowledge of reservoir pressure can also be used to estimate stress and strain variations outside the reservoir (Herwanger and Horne, 2005; Dusseault et al., 2007; Scott, 2007). Identifying those stress patterns helps guide drilling decisions and reduce the cost of repairing or replacing wells snapped or sheared by high stresses.

Conventional methodologies employ poststack data and compaction-induced vertical stress/strain to estimate time-lapse velocity and volume changes (Hatchell and Bourne, 2005; Janssen et al., 2006; Roste, 2007). However, migration and stacking of data represents a complex filtering process that can corrupt phase rela-
tionships and arrival times. Further, velocity/strain estimation from field data using this approach rarely produces results that agree with laboratory experiments (Bathija et al., 2009). Additionally, it has been shown that shear (deviatoric) strains generate significant time shifts, requiring the use of triaxial geomechanical interpretation of time-lapse data (Schutjens et al., 2004; Sayers and Schutjens, 2007; Herwanger, 2008; Sayers, 2010; Smith and Tsvankin, 2011b).

Estimation of compaction-related time shifts requires geomechanical computation of excess strains, strain-induced stiffnesses, and modeling of time-lapse wavefields. Fuck et al. (2009, 2011) develop a modeling methodology based on triaxial strain formulation and nonlinear theory of elasticity, and estimate P-wave shifts using anisotropic ray tracing. Smith and Tsvankin (2011b) confirm the P-wave results of Fuck et al. (2009) and analyze time shifts for S- and PS-waves using finite-difference elastic modeling. These studies demonstrate that both volumetric (hydrostatic) and deviatoric (shear) strains generate significant time-shift contributions for all three (P, S, and PS) modes. According to their results, sensitivity of time shifts to reservoir pressure strongly varies with wave type and reflector location.

Here, we use the geomechanical and seismic modeling methodology of Smith and Tsvankin (2011b) to study the dependence of P-, S-, and PS-wave time shifts on reservoir pressure. For a set of reflectors located above and below the reservoir, we examine the linearity of time shifts expressed as a function of reservoir pressure for different noise levels in the data. We also examine time-shift misfits with respect to a reference reservoir for a realistic range of pressure reductions. Further, we estimate time-shift variations for a range of the third-order stiffness coefficients and reservoir widths - parameters that will be assumed to be known during the inversion.

**Theoretical Background**

Modeling traveltime shifts caused by production-induced changes in a reservoir is a three-step process. First, changes in reservoir parameters (here, pressure reduction) result in excess stress and strain in and around the reservoir. Second, the excess stress/strain perturbs the stiffness coefficients \(C_{ij}\) that govern the velocities and traveltimes of seismic waves. Third, the stress-induced stiffnesses are used to model seismic data and compute the time shifts between the baseline and monitor surveys. Note that time shifts are generally non-linear in the relevant stiffness coefficients even when the model is homogeneous and isotropic (Figure 1). In the following tests, we compute time shifts for the reflectors shown in Figure 2.

**Figure 1.** Reservoir geometry after Fuck et al. (2009). Pore-pressure \(p_{\text{fluid}}\) reduction occurs only within the reservoir, resulting in an anisotropic velocity field due to the excess stress and strain. For geomechanical modeling, the reservoir is located in a model space measuring 20 km × 10 km, which is sufficient for obtaining stress, strain, and displacement close to those for a half-space. The reservoir is comprised of and embedded in homogeneous Berea sandstone \((V_p = 2300 \text{ m/s}, V_s = 1456 \text{ m/s}, \rho = 2140 \text{ kg/m}^3)\) with the following third-order stiffness coefficients: \(C_{111} = -13.904 \text{ GPa}, C_{112} = 533 \text{ GPa}, \) and \(C_{155} = 481 \text{ GPa} \) (Figure 4, sample index 6) (Prioul et al., 2004). The effective stress (Biot) coefficient \(\alpha\) for the reservoir is 0.85. Velocities in the model are reduced by 10% from the laboratory values to account for the difference between static and dynamic stiffnesses in low-porosity rocks (Yale and Jamieson, 1994).

**Figure 2.** Reservoir (shaded) and reflectors (marked A, B, and C) used in our study. Gray diamonds on each reflector mark points at which we sample volumetric and deviatoric stress/strain (Figure 3).
Strain, stiffness, and ray-based time shifts

We employ a simplified, 2D rectangular reservoir model after Fuck et al. (2009, 2011) (Figure 1), comprised of isotropic Berea sandstone that follows standard Biot-Willis compaction theory (Hofmann et al., 2005; Zoback, 2007). The effective pressure in the reservoir ($P_{\text{eff}}$) changes according to a reduction of the pore fluid pressure ($P_{\text{fluid}}$):

$$P_{\text{eff}} = P_c - \alpha P_{\text{fluid}},$$

where ($P_c = \rho gz$) is the confining pressure of the overburden, and $\alpha$ is known as the effective stress coefficient (Biot-Willis coefficient for “dry” rock, where air is the only pore infill). Pressure changes occur only in the reservoir block. The resulting displacement, stress, and strain changes throughout the section can be computed from analytic equations discussed by Hu (1989), Downs and Faux (1995), and Davies (2003). However, here we perform geomechanical modeling using the finite element, plane-strain solver from COMSOL® (COMSOL AB, 2008), which can handle more complicated models. For isotropic background material (Figure 1), COMSOL solves the linear system

$$\sigma = D\epsilon_o + \sigma_o - \Delta P,$$  

where the geomechanical model has zero initial stress and strain ($\sigma_o$, $\epsilon_o$), and $D$ are the stiffness coefficients. Typical modeled volumetric and deviatoric strains for the range of pressures used in our study are shown in Figure 3 (next page) for reflector A (Figure 2) and for a horizontal line through the vertical center of the reservoir. The modeled strains are linear in the pressure drop, and 1-2 orders of magnitude higher inside the reservoir than in the overburden.

The strain-induced variations of the stiffness tensor can be expressed using the so-called nonlinear theory of elasticity (Hearmon, 1953; Thurston and Brugger, 1964; Fuck et al., 2009):

$$c_{ijkl} = C_{ijkl}^{o} + \frac{\partial C_{ijkl}}{\partial e_{mn}} \Delta e_{mn} = C_{ijkl}^{o} + C_{ijklmn}^{o} \Delta e_{mn},$$

where $c_{ijkl}^{o}$ is the second-rank stiffness tensor of the background (unperturbed) medium, $c_{ijklmn}^{o}$ is a sixth-order tensor containing the derivatives of the second-order stiffnesses with respect to strain, and $\Delta e_{mn}$ is the excess strain. Despite the term “nonlinear,” which applies to Hooke’s law, equation 3 expresses the stiffnesses $c_{ijkl}$ as linear functions of the strain $\Delta e_{mn}$.

Wave propagation through the stressed medium may then be modeled using Hooke’s law with the stiffness tensor $c_{ijkl}$. For our 2D models, we need only two stiffnesses from the sixth-order tensor, and employ empirical values of $C_{ijklmn}^{o} = C_{oij\beta\gamma}$ measured by Sarkar et al. (2003) (the matrix $C_{oij\beta\gamma}$ is obtained by the Voigt convention (Tsankvin, 2005; Fuck and Tsankvin, 2009)). Values of the stiffnesses $C_{111}$ and $C_{112}$ for different samples of Berea sandstone are displayed in Figure 4.

Fuck et al. (2011) estimate the P-wave time shifts $\delta t$ along a certain raypath using a Fermat integral with the integrand linearized in the excess strains:

$$\delta t = \frac{1}{2} \int_{r_1}^{r_2} \left[ B_1 \Delta e_{kk} + B_2 \left( n^T \Delta \epsilon n \right) \right] d\tau,$$

where $\epsilon_{kk}$ is the volumetric strain ($1/3$ of the trace of strain tensor), $\epsilon$ is the deviatoric strain, $n$ is the slowness vector, and

$$B_1 = \frac{C_{111} + 2C_{112}}{3C_{33}}, \quad B_2 = \frac{C_{111} - C_{112}}{C_{33}}.$$  

Here, $C_{33}^{o}$ is the background stiffness coefficient, and $C_{111}$ and $C_{112}$ are the “third-order” stiffness coefficients from equation 3. Equation 4 is valid only for small stiffness perturbations. Clearly, the time shifts described by equation 4 are linear in excess strains and the pressure drop inside the reservoir. In general, however, time shifts are nonlinear in the stiffness coefficients. Indeed, even the P-wave velocity in homogeneous, isotropic media is given by $\sqrt{\frac{C_{33}}{\rho}}$, and traveltime over distance $R$ is

$$t = R \sqrt{\frac{\rho}{C_{33}}}.$$  

Therefore, the issue is the range of $\Delta e_{mn}$ and $\Delta C_{ij}$ for which traveltime shifts can be accurately described as linear functions of $C_{ij}$’s and therefore, reservoir pressure reduction.
Figure 3. Typical strains generated by plane-strain geomechanical modeling of a reservoir at 1.5 km depth (Figure 1). (a,b) Volumetric and (c,d) deviatoric strains at (a,c) reflector A and (b,d) at the depth of the reservoir (Figure 2). The strains are measured at the locations of the diamond markers on reflector A, and on a horizontal line through the reservoir center in Figure 2. Legends on each plots indicate horizontal distances from the reservoir center.

**Time-shift trends vs reflector location**

Smith and Tsvankin (2011b) employ an elastic finite-difference algorithm to compute time shifts of P-, S-, and PS- reflections for the model in Figure 1. Figure 5 shows typical time shifts for a reservoir at 1.5 km depth with a pressure drop of 20%. For P-waves (Figure 5a) the results are close to those obtained by ray tracing (Fuck et al., 2009). Strain-induced P-wave velocity anisotropy around the reservoir causes offset-dependent traveltime shifts. Laterally varying P-wave time-shift patterns below the reservoir are due to elevated shearing (deviatoric) strains at the reservoir endcaps.

Spatial distributions of time shifts show that specific wave types/reflector combinations exhibit different sensitivities to reservoir pressure changes. P-waves both above and below the reservoir exhibit distinct time-shift variations with offset due to changing deviatoric stress around the reservoir (Figure 5a). The combination of increased volumetric and deviatoric strains inside the reservoir generates large S- and PS-wave time shifts from reflectors beneath it (Figures 5b,c). These
trends provide useful guidance for designing a pressure-inversion procedure. For this study, the reflectors shown in Figure 2 are used to measure time shifts of P-, S- and PS-waves for a range of pressure drops and evaluate the linearity of the pressure dependence of time shifts.

**Dependence of time shifts on reservoir depressurization**

The character of the variation of time shifts from specific reflectors with pressure can determine the methodology used to invert for reservoir pressure changes. If time shifts change linearly with pressure, they can be represented by a system of linear equations for a multicompartment reservoir:

\[
\mathbf{A} \Delta \mathbf{P} = \Delta \mathbf{t},
\]

where \(\Delta \mathbf{P}\) is a vector containing pressure drops in all reservoir compartments, and \(\Delta \mathbf{t}\) is a vector of P-, S-, and/or PS-wave time shifts for a range of receiver offsets. Elements of matrix \(\mathbf{A}\) are the pressure-dependent time-shift gradients \((\partial t_i(x, z)/\partial P_j)\) (\(i\) denotes offset and \(j\) reservoir compartment) of reflections for specific source-receiver pairs. Therefore, for small perturbations in \(C_{ij}\), the coefficients of \(\mathbf{A}\) can be computed using standard linear inversion techniques (e.g., Gubbins, 2004) for a combination of reservoir geometry, source location, and reflector coordinates.

**Methodology**

**Modeling and assumptions**

We employ the modeling methodology described by Smith and Tsvankin (2011b) to investigate the behavior of time shifts for the reflectors shown in Figure 2. The initial reservoir pressure corresponds to a state of stress/strain equilibrium. Thus, the initially homogeneous stiffness/velocity field across the section becomes heterogeneous as reservoir pressure is reduced. Following geomechanical finite-element modeling of pressure-induced strain \(\Delta e_{mn}(x, z)\), changes in the stiffnesses are computed from equation 3. These stiffnesses are used by an elastic finite-difference code (Sava et al., 2010) to generate shot records.

Reflectors A, B, and C shown in Figure 2 are inserted in the model to sample travel times and estimate time shifts for each shot. Then, the multicomponent synthetic data are processed to isolate arrivals from the specific reflector, and time shifts between the reference (baseline) and monitor reservoir models are computed by cross-correlation. P-wave time shifts are measured from the vertical displacement, while S- and PS-shifts are measured on the horizontal component. Additional smoothing is applied to reduce interference-related distortions. For all models discussed below, the reservoir

---

**Figure 5.** Typical time-shift surfaces for (a) P-waves, (b) S-waves, and (c) PS-waves, measured using 22 reflectors around the reservoir (white box) of Figure 1 (Smith and Tsvankin, 2011b). The time shifts correspond to hypothetical specular reflection points at each (X,Z) location in the subsurface. Positive shifts indicate lags where monitor survey reflections arrive later than the baseline events; negative shifts are leads. Source location is indicated by the white asterisk at top.
depth is 1.5 km, and the source is located above the center of the reservoir at X=0 km.

**Measurements and misfits**

In the tests below, we individually vary reservoir pressure, stiffness coefficients $C_{111}$ and $C_{112}$, and reservoir width. A pressure drop of at least 30% is typical for most reservoirs, while the coefficients of $C_{111}$ and $C_{112}$ are measured with significant uncertainty (Figure 4)(Sarkar et al., 2003). Reservoir depth is usually well known, but the reservoir width may be estimated with an error. Shear (deviatoric) strains concentrate at the endcaps of the reservoir model (Smith and Tsvankin, 2011b), and changes in width will move the location of these anomalies and change the ratio of volumetric-to-deviatoric strain across the reservoir.

Misfits for P-, S-, and PS-waves time shifts between a given reservoir and a reference reservoir model are computed as the L2-norm,

$$\mu = \sqrt{\sum_{k=1}^{N} (\Delta t_{k}^{\text{ref}} - \Delta t_{k})^2}, \quad (8)$$

where $k = 1, 2, ..., N$ are the individual traces in the shot record. When computing the time-shift misfit between a modeled (test) reservoir and the reference reservoir, both reservoirs have been depressurized from a zero-stress/stress initial state. Joint misfit is the sum of the L2-norms for all wave types at a specific reflector. Misfits shown below have not been normalized in order to facilitate comparison of results for different wave types at specific reflectors.

**ANALYSIS**

**Time shifts and sensitivity to reservoir pressure**

Figure 6 shows measured time shifts of P-, S-, and PS-waves (columns) reflected from interfaces A, B, and C (rows) of Figure 2 for a set of 20 reservoir pressure drops ranging between 0.01% and 30% of the initial reservoir fluid pressure. Positive shifts indicate lags where monitor survey reflections arrive later than baseline survey reflections, while negative shifts are leads. Data for S- and PS-waves do not include time-shift estimates at X=0 due to the low amplitudes of the horizontal displacement at small offsets.

In general, P-wave time shifts at reflector A (top row) are linear lags in pressure drop, caused by a P-wave velocity reduction above the reservoir. S-waves at reflector A experience small velocity increases and time shifts due to changes in the stiffness $C_{025}$ in the overburden. PS-wave shifts above the reservoir are close to zero, because P-lags are almost canceled by S-leads (Smith and Tsvankin, 2011b). At reflector B (middle row) all time shifts exhibit slightly nonlinear behavior with increasing pressure drop. Time shifts for all wave types from reflector C are clearly nonlinear as a function of pressure because of large stiffness perturbations inside the reservoir. Therefore, we can invert for time shifts at reflectors above the reservoir using the linear system of equation 7. While reflections from beneath the reservoir may be approximated as linear for small pressure drops, they become nonlinear after 10-20% pressure reduction, which necessitates application of a nonlinear inversion algorithm.

As a second aspect of our 4D inversion feasibility study, we evaluate the sensitivity of “compound” time shifts for a certain mode and a given reflector to a range of reservoir depressurizations. For this discussion, the term “higher sensitivity” indicates a combination of high time-shift values with respect to reservoir pressure variation, and a steeply-sloped misfit curve with a distinct minimum misfit at the reference (true) pressure value (Figure 7c, for example). L2-norm time-shift misfits (equation 8) for 20 pressure drops between 0.01% and 30% were computed against a reference reservoir in Figure 1, which has a pressure drop of 15%. The results for P-, S-, PS-waves and joint misfits (equation 8) at reflectors A, B, and C (Figure 2) are shown in Figure 7. Misfit curves correlate well with the time-shift magnitudes for each wave type and reflector depth in Figures 2 and 5. At reflector A, PS-wave time shifts at larger offsets provide greater sensitivity to lower pressure drops than time shifts of P- and S-waves. At the top of the reservoir (reflector B), P-wave time shifts change most rapidly with pressure deviation from the reference value. S-wave shifts clearly provide the largest sensitivity for all pressure drops beneath the reservoir at reflector C. However, in all cases, the joint misfit is more sensitive to pressure than the misfit for any single wave type.

**Influence of the coefficients $C_{111}$ and $C_{112}$**

The third-order stiffnesses $C_{111}$ and $C_{112}$ control the magnitude of the second-rank stiffness tensor $c_{ijkl}$ computed from excess strain (equation 3). Due to large variations of $C_{111}$ and $C_{112}$ over small stress/pressure ranges (Sarkar et al., 2003; Bathija et al., 2009), scarcity of available data, and because they are part of a linearized equation, these coefficients may represent a source of significant error in time shifts. Equation 5 shows that the value of $C_{112}$ controls the relative contributions of the volumetric and deviatoric strains to time shifts in equation 4.

Variations of both $C_{111}$ and $C_{112}$ produce smaller time-shift magnitudes above the reservoir at reflector A than those for pressure variations (having a maximum of approximately 4 ms). The largest time shifts due to both coefficients occur at reflector C, ranging up to -20 ms for P-waves, -30 to -40 ms for S-waves, and -20 to -30 ms for PS-waves. Misfit sensitivity curves for variations in both coefficients are shown in Figure 8. With the exception
Compaction-induced reservoir time-shift inversion feasibility

Figure 6. Time shifts for reservoirs with pressure drops ranging from 0.01% and 30%. The source is located above the center of the reservoir at X=0. (a,b,c) reflector A, (d,e,f) reflector B, and (g,h,i) reflector C. Columns correspond to (a,d,g) P-wave, (b,e,h) S-wave, and (c,f,i) PS-wave. Plot legends indicate receiver (surface) coordinate between X=0 km and X=2.0 km.

Figure 7. L2-norm misfits of time shifts for reservoir pressure drops ranging from 0% to 30%. The reference reservoir is maintained at a pressure drop of 15%. (a) reflector A, (b) reflector B, and (c) reflector C.

Influence of reservoir width

While reservoir depth is typically well known from borehole data, the true width of the reservoir may be estimated with an error. Shear (deviatoric) stresses are largest at the endcaps of the reservoir, even for reser-
voirs of elliptical shape (Smith and Tsvankin, 2011b). The distance between these shear-strain anomalies will vary with reservoir width, thus changing the strain field near and around the reservoir.

Figure 9 shows estimated time shifts of P-, S- and PS-waves (columns) at reflectors A, B, and C of Figure 2 (rows) for reservoir width ranging from 0.5 km to 4 km, in 0.5 km intervals. The reference reservoir width for misfit measurements is 2 km. Time shifts above reflector A do not vary significantly with reservoir width, except when shearing strains from the reservoir endcaps are close to one another. However, directly above the reservoir at reflector B, P-waves time shifts change by up to 10 ms. At reflector C, maximum time shifts occur for larger widths, and are similar to those of the reservoir with ΔP = 20% shown in Figure 5. L2-norm sensitivity curves for P-, S-, and PS-wave data shown in Figure 10 are reasonably smooth, with the exception of S-wave misfit on reflector B for a reservoir width of 3 km. Joint misfit data from reflector C are most sensitive to variations in reservoir width.

Sensitivity to noise in time shifts
The influence of Gaussian noise with standard deviations of 2, 5, and 10 ms added to the reference time shifts (Figures 6 and 7) is shown in Figure 11. Time-shift misfits for 2-ms noise (left column) do not differ significantly from the corresponding noise-free estimates in Figure 7. Substantial degradation in sensitivity is observed for reflectors A and B (above reservoir) for 5-ms noise (Figures 11b,e). The misfit curves develop local minima, indicating that a linear inversion algorithm may fail at moderate noise levels. For strong noise reaching 10 ms (approximately 2/3 of the maximum P-wave time shifts for noise-free data), the misfit curves for reflectors A and B are significantly distorted. However, time shifts of all wave types for reflector C are sufficiently large to provide smooth sensitivity curves with a clear global minimum and minimal degradation. In general, as noise levels increase, the sensitivity of individual wave types/reflectors to pressure reduction declines, but joint sensitivity at reflectors B and C remains reasonably high. These results suggest that in the presence of substantial noise, joint and S-wave time shifts for reflectors below the reservoir provide the most reliable input data for pressure inversion.

Conclusions
We have used a simple, 2D rectangular reservoir model to study the dependence of P-, S- and PS-wave time shifts on reservoir pressure with the goal of assessing the feasibility of pressure estimation. The model is comprised of a single homogeneous block of Berea sand-
Compaction-induced reservoir time-shift inversion feasibility

Figure 9. Time shifts for reservoirs with widths ranging from 0.5 km to 4 km. Pressure drops for all reservoirs are 15%. (a,b,c) reflector A, (d,e,f) reflector B, and (g,h,i) reflector C. (a,d,g) P-wave, (b,e,h) S-wave, and (c,f,i) PS-wave. Plot legends indicate receiver (surface) coordinate between X=0 km and X=2.0 km.

Figure 10. L2-norm misfits of time-shift data for reservoir widths ranging from 0.5 km to 4 km computed with respect to a 2 km-wide reference reservoir (Figure 1). The pressure drop for all reservoirs is 15%. (a) Reflector A, (b) reflector B, and (c) reflector C (see Figure 2).

stone, where pore pressure changes inside the designated reservoir block induce heterogeneous stress/strain and stiffness fields throughout the medium. Geomechanical modeling is implemented with a finite-element solver that generates excess stress and strain as a linear function of reservoir pressure. Multicomponent seismic data are modeled by an elastic finite-difference code, and resulting time shifts are computed by specialized post-processing. While the stress-induced stiffness tensor is
linear in excess strain, traveltime shifts generally depend
on the stiffness coefficients in a nonlinear fashion.

In the regions with relatively small strain, pressure-
related perturbations in the stiffnesses are not suffi-
ciently large to cause nonlinearity of time shifts. For ex-
ample, time shifts are linear in pressure reduction for re-

deracter A above the reservoir. However, strains inside the
reservoir are much larger than those in the surrounding
medium, creating large stiffness changes. Thus, waves
reflected from points at and below the reservoir exhibit
nonlinear time-shift dependence on pressure.

L2-norm misfits of time shifts computed with re-
spect to a reference reservoir show that S-wave time
shifts from reflectors located beneath the reservoir are
generally most sensitive to pressure. Misfit curves for
S-wave shifts from deep reflectors provide robust indi-
cators of pressure change even for reference time shifts
contaminated with 10-ms noise. Therefore, inversions
for reservoir pressure with noisy data need to operate
with reflections from beneath the reservoir. Variations of
the third-order stiffness coefficients ($C_{111}$ and $C_{112}$) and
reservoir width perturb the stress/strain fields around
the reservoir, but do not cause distortions that could
seriously impede pressure inversion.

Based on our findings, for cases of small pressure
changes, a linear inversion method could be sufficient
for pressure estimation. However, nonlinear time shifts
at pressure drops above 10% should be inverted by a
global algorithm.

Ongoing work using the modeling methods dis-

cussed here includes the investigation of multicompart-
ment reservoirs with interacting stress and strain fields.
To estimate reservoir pressure, we are devising a non-
linear global/gradient-hybrid inversion based on the
nearest-neighbor algorithm of Sambridge (1999). After
testing on the simple reservoir model discussed here,
this algorithm will be extended to multicompartment
reservoirs.

Figure 11. L2-norm misfits for noise-contaminated reference time shifts as a function of reservoir pressure drop (compare with
corresponding noise-free misfits in Figure 7 ). Reference reservoir pressure drop is 15%. (a,b,c) reflector A, (d,e,f) reflector B,
and (g,h,i) reflector C. Standard deviation of noise in time shifts increases by column: (a,d,g) ±2 ms, (b,e,h) ±5 ms, and (c,f,i) ±10 ms.
Acknowledgments

We are grateful to Rodrigo Fuck (HRT-Brazil), Mike Batzle (CSM/CRA), Roel Snieder (CWP), and Jay Behura (CWP) for discussions, comments, and suggestions, and to Jeff Godwin (CWP, now Transform Software) and John Stockwell (CWP) for technical assistance. This work was supported by the Consortium Project on Seismic Inverse Methods for Complex Structures at CWP.

References

Hatchell, P., and S. Bourne, 2005, Rocks under strain: Strain-induced time-lapse time shifts are observed for depleting reservoirs: The Leading Edge, 24, no. 12, 1222–1225.
Herwanger, J., 2008, R we there yet?: 70th Annual International Meeting, EAGE, Expanded Abstracts.
Hofmann, R., X. Xu, M. Batzle, M. Prasad, A.-K. Furre, and A. Pillitteri, 2005, Effective pressure or what is the effect of pressure?: The Leading Edge, 24, no. 12, 1256–1260.
——–, 2011b, Estimation and analysis of compaction-induced traveltime shifts for multicomponent seismic data: Submitted to Geophysics.
Teaching graduate students The Art of Science

Roel Snieder\(^{(1)}\), Ken Larner\(^{(1)}\), and Tom Boyd\(^{(2)}\)

\(^{(1)}\) Center for Wave Phenomena, Colorado School of Mines, Golden CO 80401, USA, email rsnieder@mines.edu
\(^{(2)}\) Dean of Graduate Studies, Colorado School of Mines, Golden CO 80401, USA

ABSTRACT
Graduate students traditionally learn the trade of research by working under the supervision of an advisor, much as in the medieval practice of apprenticeship. In practice, however, this model generally falls short in teaching students the broad professional skills needed to be a well-rounded researcher. While a large majority of graduate students considers professional training to be of great relevance, most graduate programs focus exclusively on disciplinary training as opposed to skills such as written and oral communication, conflict resolution, leadership, performing literature searches, teamwork, ethics, and client-interaction. Over the past decade, we have developed and taught the graduate course The Art of Science, which addresses such topics; we summarize the topics covered in the course here. In order coordinate development of professional training, the Center for Professional Education has been founded at the Colorado School of Mines. After giving an overview of the Center’s program, we sketch the challenges and opportunities in offering professional education to graduate students. Offering professional education helps create better-prepared graduates. We owe it to our students to provide them with such preparation.

Key words: education, professional development

1 INTRODUCTION, THE NEED FOR PROFESSIONAL TRAINING

Graduate school is aimed at preparing scientists and engineers to be the scientific professionals of the future. Those graduating, with either MSc or PhD degrees, pursue a career in a variety of areas of employment – in academia, industry, or government. Unsurprisingly, graduates with such advanced degrees ultimately attain high-level positions with their employers. Sometimes such position are in the research field in which students have studied, but often the professional life sooner or later extends beyond the specialist research in which the students are trained (Golde and Dore, 2001). Even those pursuing a continuation of their research after graduation usually experience a change in their daily activities as they acquire progressively more responsibility such as in leadership of research groups. This suggests that graduate programs should prepare students to assume broad professional roles that go beyond those of the specialized individual researcher: universities aim to educate the professionals of the future and should educate students accordingly.

One might question the extent to which graduate students are, in practice, adequately prepared for carrying out their research while in graduate school. Where do students, for example, learn how to successfully choose a research topic, prepare a research plan, work effectively with an advisor, do a literature search and archive results, manage their time, and communicate effectively? Whereas one might hope that students learn such skills from their advisors and by being part of a research group, but how well are they in fact learning such skills in this way?

The mechanism for educating graduate students is essentially a medieval system wherein the pupil (the student) works with the master (the advisor) for several years. Once the master decides that the pupil has learned her new trade, the time has come to award the graduate degree. This is, of course, a caricature of graduate school, but it is one with a grain of truth. In practice, while students focus on research and disciplinary classes, some of them receive frequent and excellent coaching from advisors and team-members. Many graduate students, however, lack adequate mentoring in professional skills. This often leads to a loss of time in graduate school, needless frustration and discouragement, and an
overly narrow preparation for a future career. As stated by Cassuto (2011)

“It amounts to this: Graduate school is professional school, but most PhD programs badly neglect graduate students’ professional development. We spend years of their training ignoring that development, and then, only at the last moment when students are about to hit the job market, do we attend to their immediate professional needs.”

This view of graduate school might, in fact, be overly optimistic since it assumes that graduate students who are near the end of graduate training are actually taught professional skills. That might not even be the case.

The National Institutes of Health (NIH) carried out a study to determine the effectiveness of professional training given to graduate students and postdoctoral fellows supported by NIH (Mervis, 2011; NIH, 2011). The report identifies the need for adequate professional training. Berg (NIH, 2011) writes in his role as director of the National Institute of General Medical Science:

“Ultimately, a healthy biomedical and behavioral research enterprise requires that government, academia, industry and other partners work together toward common goals that recognize the essentiality of high-quality mentoring and career guidance for the next generation of scientists. Our future, the future of discovery, and the utilization of such discovery for the benefit of humankind depend on it.”

This quote expresses that adequate mentoring and training of young researchers goes beyond the interest of individual researchers; the effectiveness of the whole research endeavor and its potential impact on society might depend on the degree to which young researchers are adequately prepared beyond the confines of their specific scientific field. Clearly there is a need to focus on the professional education of young researchers.

2 ARE GRADUATE STUDENTS BEING ADEQUATELY TRAINED?

Few systematic studies have been conducted that quantify to what extent the needs of graduate students for professional training are met. The American Institute of Physics carried out an “initial employment survey” among physics PhDs with first jobs in the private sector who graduated in 2007 or 2008 (Ivie, 2011). This survey showed that 98% of the graduates stated they work with their new employers on teams, 67% regularly speak in public, and 63% work with clients. This raises the question “Have these students been adequately trained in graduate school in the ‘soft’ areas such as team-work, oral communication, and client interaction?”

We asked a mix of MSc and PhD students in different stages of their graduate studies their views about the value of a range of professional skills including communication, conflict resolution, leadership, carrying out literature surveys, and teamwork. A large majority of the graduate students acknowledged the importance of broad professional training in general, e.g., written and oral communication, ethics, and leadership. Moreover, employers consider such skills to be essential to advancement of their young professionals and, by extension, to the success of their organization.

In our roles of advisor and dean of graduate studies, we discovered that the adaptation of students from different cultures to the work style in the USA can take a considerable effort of the student and the advisor. Interestingly, NIH identifies the creation of a diverse workforce as one of five priority items; the NIH-report (2011) states that “diversity is an indispensable component of research training excellence, and it must be advanced across the entire research enterprise.” We are currently developing a 1-credit graduate course that helps students better understand cross-cultural issues that can arise in graduate school.

We also asked graduate students what they believe to be the most efficient method for delivering professional training. Surprisingly, students thought that interaction with advisors was the least effective way to provide such training. This is in striking contrast with the notion that the advisor is the person most important in helping their students grow to become young professionals. This suggests that those graduate students have not perceived the mentoring they experience to be of particular value to their professional growth. This does not mean, of course, that mentoring and advising is unimportant; rather, it reflects the fact that interactions with advisors currently often fall short of inspiring acquisition of general professional skills. Coaching academic faculty – training the trainers – might have fundamental value for increasing the effectiveness of academic advising and benefitting the professional growth of young researchers.

To advance their professional skills, students have the greatest preference for seminars or workshops, or to incorporate elements of professional training into regular courses. These two approaches are perceived to require their smallest investment of time, and therefore represent for them the most sensible options given the pressure they are under to finish a graduate degree in a limited amount of time. This preference for time-efficient options is, in practice, reinforced by some advisors who question the merit of broad stand-alone professional training. This attitude among academic advisors might be a result of their not having received formal professional education themselves.

Incorporating elements of professional education into disciplinary courses is not only time-efficient, it also has the merit of placing professional skills in the context of the discipline chosen by the student. One might question, however, how easy this option is to realize in practice because many disciplinary courses are already overloaded with content judged most pertinent to advancing students’ research; moreover, teachers of-
ten have neither the interest nor the skills to broaden the scope of their disciplinary classes to include aspects of professional education.

In order to help students develop effective research habits and grow into professional researchers, in 2002 one of us (Roel Snieder) developed the course *The Art of Science* at the Colorado School of Mines (CSM). In the next section we give an overview of topics covered in that course. Despite its relatively small size, CSM now offers six different graduate courses for professional training of graduate students. Furthermore, in order to coordinate these courses and to foster new initiatives for professional education of graduate students, CSM has founded a Center for Professional Education. We describe the scope of this center in more detail in section 4.

3 THE ART OF SCIENCE

The course *The Art of Science* started at a modest scale, with about five students in the first year. Currently the class attracts about 60 students per year, which is about 25% of the graduate students that enter CSM each year. The class is taught to students from all departments. Several departments have made this graduate course mandatory for their students because they have found students to be more effective in their research after they have taken this class.

The growth in the enrollment in the course might be attributed to a combination of factors; one is that the class, which started out as a 3-credit course, was subsequently shortened to being a 1-credit one. This reduction in time-investment for the course was appreciated by students, and it also made advisors more receptive to their students ‘sacrificing’ some research time to take this class. This reinforces the conclusion of section 2 that it is important that professional education for graduate students be offered in a manner that is time-efficient.

3.1 Course content

The curriculum of the class, including its homework exercises, can be accessed online.* The course currently covers the following topics, that range from the philosophical to the rather nuts-and-bolts applied:

- What is science?
- Making choices
- The advisor and thesis committee
- Questions drive research
- Giving direction to your work
- Turning challenges into opportunities
- Ethics of research
- Using the scientific literature

*http://inside.mines.edu/~rsnieder/Art_of_Science.html

![Figure 1. Cover of the book The Art of Being a Scientist.](image)

Each of these topics is covered in much more detail in the textbook (Snieder and Larner, 2009) (see Figure 1) that grew out of this class; here we give just an overview of each of the topics in the course.

What is science?

Given the breadth and depth of the topic of the philosophy of science, one could readily teach a separate full course on the scientific method. For students who are almost always pressed for time, however, it is useful to restrict the class to material that has immediate and practical implications for their research. Science is based on logic; a statement that is inconsistent with logic might well be interesting or beautiful, but it cannot be considered part of science. Science nevertheless often makes its largest advances based on such non-logical, ill-defined abilities as creativity, inspiration, insight, and intuition, and the seemingly unscientific activity of play. And, then there’s the fortuitous element – serendipidity. Successful scientists appreciate the great paradox that science is an activity that, while a description of nature dependent on logic, often moves forward through pathways that are not logical at all! Although students are
Making choices

Students face many decisions in their choice of graduate program, advisor, research topic, and future career path. The need to make choices, of course, continues after graduate school as well; whenever confronted with having to select what is hoped will be a good decision, it is essential to be well informed. One part of being informed is to get the information from the right people. When choosing an advisor, for example, it is not only important to talk with potential advisors, it also pays off to talk with their students, especially former students, as well. Making the right choice of research topic has far-reaching implications for success in graduate school and satisfaction in later career and life. A suitable research topic must be innovative and doable, it must match the research facilities and the time that are available, and it should also offer the promise of instilling a passion in those doing the research. We teach students the concept of the “S-curve of development” wherein a research field goes from initiation stage through exponential growth to maturation. Having (1) an awareness of differences in challenges and opportunities that these three stages present, (2) a recognition of the stage a particular line of research is currently in or might soon enough be in, and (3) an understanding of which stage of research best matches the student’s technical, creative, and emotional strengths can assist in making the right choice of research topic.

The advisor and thesis committee

The advisor is a central person in the career of a graduate student. Not only does the advisor need to sign off on specific stages of graduate study, of more consequence is the essential role that the advisor ideally plays in mentoring the graduate student towards scientific independence. The perfect advisor is a creative and respected researcher, a dedicated mentor (Vesilind, 2001), she challenges the student but provides support when needed, has ample time, and has financial resources and infrastructure for the research. Clearly that perfect advisor is hard to find. The situation is, in fact, even more complicated. Different students need different types of advisors. The insecure over-achiever needs, for example, a different style of supervision than does an overconfident self-starter. As students change over time – and they should change as they grow through the graduate program – the style of advising that works best for them will also change. Given that the ‘perfect advisor’ is an illusion, a student’s task is to try to make an optimal choice, one that likely involves some degree of compromise, rather than a perfect choice. In The Art of Science, we aim to help students by making them aware of the elements that could be part of the choice and by giving them ideas for going through the process of choosing an advisor. Students, moreover, need to learn how to make effective use of their advisor (Kearns and Gardiner, 2011). Related to this is their use of the thesis committee. Many graduate students see this committee just as a machine to provide signatures, but with the right outlook (of both the student and the committee), the committee can play a valuable role as sounding board for crucial choices in a student’s research and education.

Questions drive research

An interesting exercise is to pose graduate students the following assignment. Please complete the following sentence in not more than 20 words: “The main question I want to address in my research is ...” It might be surprising to find that many students have difficulty in completing this sentence. That difficulty arises because they lack clarity on the very research question they aim to address. But how can one expect to find an answer if it is not clear what the question is? It is fundamental that students understand the significance of asking questions in research. Some of these questions are major overlying ones, some are more specific with a focus on a particular practical problem. The ‘right’ questions are ones that almost automatically lead to actions in research. The best start at asking the ‘right’ question comes from asking lots of them – both ‘good’ ones and ‘stupid’ ones, simple ones and bold ones. It can take courage and imagination to ask bold questions. The question “What would be the consequences if the speed of light were the same for all observers?” posed by Einstein is at first sight nonsensical, yet it changed our world-view because it led to the theory of special relativity. In the course, we offer students ideas for generating research questions, ideas that range from writing them down as they arise during the day (or night), to talking with others (students don’t often enough recognize one another as a resource), and to free association. The questions thus generated form a natural basis for making a work-plan for the research project.

Giving direction to your work

A quote from Lewis Carroll states that “If you don’t know where you are going, any road will take you there.” The purpose of setting goals is to clearly define where
you are heading and what you want to achieve. Without articulating goals, random events or other people are likely to define the way in which our life and career unfold. Students should learn that if they don’t define where they are headed, somebody or something else will (indeed, also should researchers and people in general). But setting goals is not all; one also needs to be aware of process. What is the value of reaching a goal in graduate studies when the process of getting there is not attractive? Clearly the process of doing research must have its rewards as well. This offers a second perspective on giving direction to our work: being process-oriented. Ultimately a third perspective can come to the fore: what is the meaning of our work? What does our work mean to us, to the scientific community, to other people, and to the world? Whose life is touched or improved by my work? The key in setting direction is to reconcile goals, process, and meaning. This is not easy, and many of us never get to that point, but being aware of these complementary aspects of our work is an indispensable first step.

Turning challenges into opportunities

Research is challenging. One challenge that is bound to arise – often – is being stuck. There are, roughly speaking, two reasons for being stuck. The first is having insufficient clarity on the research question that one aims to address or on the path that one intends to take to address those questions. This problem can be paralyzing and needs to be fixed as soon as possible using the techniques described in the section on posing questions. The second reason for being stuck is that something is ‘wrong’ in the sense that one’s understanding is incomplete. While frustrating, this source can actually be highly positive because it can be a precursor to gaining new insights (Kuhn, 1962). Not a small aspect of his approach to research, Lord Kelvin emphasized “When you are face-to-face with a difficulty, you are up against a discovery.” Students need to learn what activities help them in getting unstuck. For some of us, for example, running is an activity that is conducive for getting new insights when these are needed. By developing an awareness of activities that are helpful in getting unstuck, students can become more effective in getting out of seeming impasses they invariably will encounter. In the course, we also discuss the role that play and serendipity have in opening new avenues of thought and for research.

Ethics of research

The ethics of research is often seen in the restricted sense of “responsible conduct in research.” That view of research ethics focuses on issues such as honesty, avoiding plagiarism, and appropriate sharing of authorship – important issues that students need be aware of. In practice one often cannot repeat and verify experiments done by others nor can one easily chase down all possible antecedents to reported research. Students therefore must recognize the extent to which the advancement of science is founded on trust. Moreover, students need to be made aware that honesty and preservation of one’s reputation is especially challenged when economic interest lies at the heart of a scientific or engineering endeavor. Yet there is more to research ethics than the need for honesty. It is through dissent that science moves forward beyond current understanding; science cannot advance when everyone is in agreement and is satisfied with current viewpoints. In his wonderful book “Science and Human Values,” Bronowski (1956) emphasizes that science is underpinned by honesty and advanced by dissent. Yet, as illustrated in the parody of Oxman et al. (2004), not every scientist has developed the skills to disagree respectfully and express dissent in a constructive way. Apart from these issues of how research is carried out, it is essential to the growth of a student into a mature and contributing researcher to gain the insight that research is not value-free. In this highly technological age, science has the potential to affect society and the lives of others either negatively or positively; often a given scientific advance can have both positive and negative ramifications. A heightened awareness of such potential consequences can be a valuable guide toward choices in research whose impact on society and its environment is positive.

Using the scientific literature

Research projects should begin with a solid literature search. Students sometimes skip or minimize this step and, as a result, either duplicate work of others or make their work unnecessarily difficult by failing to take advantage of the insights of others. Over the years, the scientific literature has grown so much that it is difficult to keep pace with. Students need to learn how to use the powerful electronic tools available for searching the literature effectively. Furthermore, it is important to develop a modern data base of references: such a database makes it possible to retrieve papers easily and to generate a bibliography at the end of papers or a thesis with the touch of a few buttons. In the course, we make our students aware of free database programs such as Zotero, Mendeley, and Jabref. As a homework exercise, students must choose a system for managing bibliographic information and show that they have started using the program of their choice. We have found that without the reinforcement of such an exercise, although the class tends to be interesting to students, it might
not lead to helpful behavioral change. Nevertheless, not all approaches for keeping up with the literature ought to rely solely on electronic tools and information technology. A journal club, for example, with stimulating discussions on research among colleagues, can be a great aid to maintaining awareness of current research.

Communication

Career advancement for a scientist, whether working in academia, industry, or government, can be governed largely by the young researcher’s ability to communicate, both orally and in writing. We emphasize that, effective communication is difficult. For both oral and written communication of research, perhaps the most difficult skill to learn is to place oneself in the shoes of the reader or listener. This skill is essential for setting the level and tone of the communication needed to reach the audience and peak its interest. In oral presentations, one must be aware of the short attention span of most audiences (Medina, 2008), indeed the relatively short time available to convey research and results that could well have taken a year or more for the presenter to have generated and understood. Young researchers need to resist the temptation of over-feeding the audience with material that, despite the availability of modern projection resources, is poorly readable (Benka, 2008; Payne and Larner, 2008). An effective way we’ve found for teaching students the do’s and don’ts of oral presentation is to show them a spoof presentation in which the teacher does everything wrong, and to discuss what is amiss in this presentation afterwards. Because few of us can give an outstanding presentation while improvising, students need to expect that preparing for an effective oral presentation is hard work that requires extensive rehearsals, the audiences for the rehearsals consisting of both colleagues who familiar with the work and those who are not. Moreover, the friendliest of rehearsal audiences has members who care enough about the success of the final product to offer in-depth critiques of both the content and details of presentation style. Similarly, any well-written manuscript usually has gone through many revisions that often rely on the input from caring colleagues who act as severe proofreaders, in essence the first line of reviewers. Also, before embarking on the writing, both students and advisors can save much time by decoupling the content of a paper from its style. This can be done effectively by first developing an outline that is so detailed that the content and flow of the paper are virtually predetermined. Once this is done, the ‘only’ thing that needs to be done is to formulate this content into words.

Our next advice is to ‘blast away’ in putting words on paper, with little concern for either the audience or word choice. Then comes the essential next step: revise, revise, revise – progressively choosing wording and writing style aimed at increasing clarity for author’s intended reader.

Publishing a paper

Obvious as it is that publishing of a paper starts with the choice of a journal, junior graduate students are often unaware of the differences in scope and quality of journals. We discuss the impact, so to speak, of the impact factor as well as other considerations to take into account when choosing a journal. Examples include the readership of the journal, the speed of publication, the cost of publication, and the journal’s reputation. As mentioned in the previous section on communication, when planning a paper for publication it is essential to know the audience for whom a paper has been written. Students should understand the mechanics of the review process and be aware of opportunities to steer that process, for example by suggesting names of reviewers and associate editor who might be best suited to handle the manuscript, people who are especially knowledgeable about the subject of the paper and who can be counted on to treat the work fairly. Also, knowing how best to respond to reviews, particularly those that might seem at first unduly critical, can be an art in itself. Students need to learn to take the comments of reviewers seriously, without being carried away by indignation when a review at first sight feels unduly critical. The comments of reviewers more often than not are a great help in improving a manuscript; they can point out segments of a paper that are in need of further clarification not only for them but for the intended readership. Students, however, also need to know that they, as authors, are allowed to deviate from the suggestions of reviewers if they can convey their reasons for disagreement. Always, the author should respond to specific concerns raised by reviewers in a cover letter to the editor, stating point-by-point how they have addressed the concerns, including those about which the author was not in agreement with the reviewer and therefore were not incorporated in the revised version of a manuscript.

Time management

Many of us feel as though there is not enough time for all that we have to do. This is, however, an illusion because, for everyone, a day always has 24 hours, and a week seven days. Nothing will change these basic facts. The feeling of not having enough time results from our trying to do too many things in a given amount of time.
The word ‘time-management’ therefore is a misnomer; the difficulty actually boils down to one of ‘activity-management’. This is not just a semantic distinction; students need to learn that is essential to make choices about which activities to spend their time on, and which not. It is illuminating for students to consider the diagram in figure 2, adapted from the book of Covey (1990). Roughly speaking, all of our activities are either important or unimportant, and they are either urgent or not urgent. This is, of course, an oversimplification because both importance and urgency vary on a sliding scale, but for the moment we make this simple distinction. Given the criteria of urgency and importance, each of our activities fits into one of the four quadrants of figure 2. Many of us make the mistake of confusing urgency with importance with the result that the important activities tend to fall by the wayside, to urgent, but actually less important, ones. As a homework exercise, students monitor their activities for a week and insert into the diagram of figure 2 each activity, with the time spent on it. Students also insert into the diagram activities that they would have liked to do, but did not find time for. After that, they analyze the way in which they have spent their time, and make a plan for improvement, if needed. We discuss in class that saying “no” is essential in activity management, and discuss why this often is so difficult. Learning how to gracefully say “no” is an important skill, and we give suggestions for ways in which to offer alternatives to requests, or directives, from superiors. (When we say “yes” to a request, we are often saying “no” to something else that we might otherwise do.) We also point out that, useful as they often are, electronic tools such as cellphones, email, and the internet, they can also be a severe distraction. We caution students to be the master of these tools, rather than their slave.

Figure 2. A categorization of activities.

The scientific career

In order to make the right choices for one’s career (choices again), the key, as before, is to be informed. We discuss the structure of the academic career and its opportunities and hurdles, such as the tenure process. Students should know that the nature of the tenure process varies significantly among universities and that the degree of fairness of the tenure process varies much as well. A useful and practical guide for best practices in tenure evaluation is available online. The choice of employment in academia versus industry or government is a major one for graduates. Because their training takes place at universities, many students in science have apparent familiarity with the university work environment, but limited understanding of what it means to work in industry or government. In the class we aim to give them insight into differences and similarities that might be expected. As a homework exercise, students interview four researchers, preferably from different types of employers, with the goal of gaining greater insight into the choices available to them. We also discuss gender issues. Despite many efforts, there still exists a gender disparity in engineering and in the higher ranks of...

www.acenet.edu/bookstore/pdf/tenure-evaluation.pdf

Writing proposals

Regardless of whether a scientists or engineer works in academia, industry, or government, writing proposals is an integral part of her job because one must, in general, explain why time and resources need to be expended on research. Every proposal starts with the choice of to whom the proposal is to be submitted, and one needs to be informed in order to make this choice wisely. When writing a proposal it is essential to stick to the guidelines given by the funding agency. Such guidelines include tangible issues such as the length of the proposal, as well as less tangible ones such as the specific points that must be addressed in the proposal. One should also be aware of the way in which proposals are handled. Many junior researchers do not realize that members of a review panel often are faced with having to read more than 1,000 pages of proposals. These panel members obviously don’t have the time to do this; hence a proposal, starting with the abstract, must make a favorable first impression. As scientists we often feel the urge to be complete in our proposals, and explain all facts and details that we feel are important to us. This, however, is not what reviewers, panel members, and program managers are looking for. They seek a concise description of the state of research in a field, the research question that one aims to address in order to advance the research field, the research methodology used, the competency and facilities of the research team, and the proposed time-line and deliverables.
scientists. We point out several mechanisms that have brought about gender bias. (Did you find the use of the word “she” instead of “he” in this chapter strange, or even disturbing?) Balancing professional and personal life is a challenge in the scientific career; we find that students are keen to discuss this topic. For female students the topic of combining motherhood and family responsibilities with a successful career is of particular interest.

**Applying for a job**

The application for a job begins with identifying several potential employment opportunities. As with all choices, one first must gather facts. For this, students should rely not only on recruiters and human resource managers, but also hear the opinion of employees and former employees. An appropriate and well-constructed letter and curriculum vitae are essential, and job seekers need to keep in mind that the relevant information must be made easily accessible for overloaded search-committee members and managers. During an interview, the person seeking employment should be pro-active. This can make the interview more useful as a fact-finding tool; moreover, a pro-active attitude usually is viewed favorably by people conducting interviews. Also, because promotions and significant salary increases are infrequent events once an individual is employed, it is important to negotiate before accepting a job offer. Although the class includes no time to cover negotiation in depth, we do point out the different styles of negotiation: win-win, win-lose, and (hopefully never) lose-lose. The job application process should aim for making an optimal match between employer and employee, which obviously calls for seeking a win-win strategy. Some industrial employers are unduly restrictive in the rights granted to their employees, for example by insisting on long-term non-compete agreements after employees leave the organization to work elsewhere. In order to avoid unpleasant surprises when starting at a new job, the job applicant should be clear about all conditions of appointment before accepting a position (Larner, 2002).

**3.2 Teaching The Art of Science**

Since 2002, we have offered *The Art of Science* as a graduate course at CSM, and have also given this class as a short-course at numerous universities that include Stanford University, Tohoku University (Sendai, Japan), Delft University of Technology (Delft, Netherlands), Australian National University (Canberra, Australia), and King Abdullah University of Science and Technology (Jeddah, Saudi Arabia, figure 3). We have also presented *The Art of Science* as a short-course for the research laboratories of ExxonMobil, Saudi-Aramco, and Shell.**

The class has consistently received positive reviews from students. One student commented that

*The Art of Science was an eye opener for me. It made me think of my career and my life differently. It gave me energy and ideas to restart and continue when I am stuck.*

When teaching physics courses, we never had a student say that our course changed their view of career and life! Having a far-reaching impact on students is an important aspect of offering a class such as *The Art of Science*. An anonymous student at an international university wrote in an evaluation that

*I am glad I found this course early in my academic career. If only my university had required faculty members to come to your class! Thank you for putting all the things together which otherwise probably would have taken me years and many unfortunate incidents to figure out.*

This comment expresses that taking a class such as *The Art of Science* can save graduate students much time. This feedback recurs often in evaluations. Students often express regret that they had not taken the class earlier because that might have saved them time by being more efficient and by avoiding time-consuming mistakes. Taking a 1-credit class in the practice of science does take some time out of a busy schedule, but the increased efficiency and effectiveness in doing research, and in communicating that research, can readily make up for the time investment. The student comment above also indicates the wish that faculty members would attend course. This points to a need for training faculty to teach the skills needed for being an effective scientist. Although most faculty members are dedicated to advance research together with their students, much could be gained by training academic faculty in mentoring.

**Please contact Roel Snieder (rsnieder@mines.edu) if you are interested in this short course.
In numerous lectures about teaching The Art of Science, we typically have received a response from faculty along the lines “it would be great to offer a course like this in our department.” The reality, though, is that developing and offering such a class takes time of faculty who are already struggling with their workload. We aim to reduce the time needed to start teaching a course such as the Art of Science by making our curriculum and homework exercises available, as examples, in the book (Snieder and Larner, 2009) and through the internet.†† Following are several options for offering this type of graduate education.

- The most straightforward scenario is to offer the class as either a departmental or interdepartmental course. This option requires a dedicated faculty member who is able and willing to champion such a course. For this scenario to work, the department or institution must recognize the value of such an educational initiative.
- One can broaden disciplinary courses to include elements of professional training. We have found that this option actually is most preferred by students who have not taken the 1-credit course. It does, however, require dedication of teachers to make time available to include professional training in their disciplinary courses; not every teacher has the skills nor perceives the time available in her course to offer such training.
- It is possible to share the teaching load by offering the class in the form of a reading group or seminar that is led, in turn, by different faculty members. This reduces the workload for individual faculty members and it might help create a greater involvement from faculty members. This scenario also makes it possible to draw upon the strengths of different faculty members.

Offering such training in any of these forms requires time of students and teachers. We do believe, however, that because of the improved efficiency that students gain, it helps them ultimately to save time. Perhaps more important, it helps students to become more creative and more efficient researchers.

4 THE CENTER FOR PROFESSIONAL EDUCATION

Currently the graduate program of the Colorado School of Mines includes the following courses for professional development:

(i) The Art of Science
(ii) Introduction to Research Ethics
(iii) College Teaching
(iv) Advanced Science Communication
(v) Academic Publishing
(vi) Professional Oral Communication

Given the small size of the school, the breadth of this course-offering speaks to the dedication of the school to professional training of graduate students. The class “Introduction to Research Ethics” was developed in response to the requirement of the National Science Foundation (NSF) that undergraduate students, graduate students, and postdoctoral fellows, receive training in research ethics.

In order to coordinate and facilitate professional education we have founded the Center for Professional Education‡‡ at CSM. Initially, the Center serves primarily graduate students, but over time might extend its activities to undergraduate education as well. The Center coordinates educational activities that include courses for broad professional development of students, seminars and workshops for students and faculty, and a speaker series. The Center brings together faculty dedicated to educating graduate students who are well prepared for the workforce, and acts as a nucleus for writing proposals to support initiatives for professional education, including new methods of delivery. Activities of the Center are directed not only toward helping graduate students, but also toward providing assistance and support to faculty so they might improve their advising skills. The Center can serve a number of purposes.

1. Develop more rounded graduates. Graduates compete with highly qualified graduates from other institutions for positions in academia, industry, and government. As such, graduates strive to find ways to distinguish themselves from the competition. While technical competence remains the most valued aspect of graduates, students can distinguish themselves through development of professional skills.

2. Advertise educational activities. The presence of the Center makes it possible to advertise our activities for professional education. This helps, for example, in recruiting top-level graduate students and in soliciting external support through foundations for graduate education and research contracts. Advertising the possibility of gaining professional skills could be particularly helpful in attracting top-level international students.

3. Support new proposals. Pressure has been increasing for generating proposals to the NSF, NIH, and other funding agencies that require broader education for graduate students. The presence of an active Center with a broad offering of courses strengthens proposals that must include elements of professional education.

4. Conduct research on education in professional development. The Center will engage in education research in professional development by initiating and coordinating such a research effort, and by helping to solicit funding for such research activities.

5. Ease the task of advising students. Much of the time spent advising students tends to be devoted to disciplinary discussions that are at the core of the research experience. Training graduate students in professional development helps them be more effective in their research, in interacting with their advisor, in improving their interaction with their advisor, and in developing essential oral and written communication skills. Such training makes the task of advising easier, and hence reduces the workload of advisors.

6. Initiate new activities and bring the relevant faculty together. The Center is a nucleus to bring together faculty with a passion for professional development and to coordinate their efforts toward the creation of new initiatives and improvements of existing efforts. The current course offering does not cover all areas in professional development that are relevant for graduate students. By organizing seminars, and workshops, and by initiating the development of new courses, the Center serves to expand the scope of professional education.

5 CONCLUSION, CHALLENGES AND ADVANTAGES OF PROFESSIONAL GRADUATE TRAINING

In offering professional education to graduate students, we have encountered a number of challenges. First, teaching aimed at encouraging change in the behavior of students to make them more effective scientists, does not necessarily lead to such behavioral change. Most teachers know the phenomenon that students who have learned certain material in class are unable to use that material a semester later. This also holds for professional training. In order for material to stick with students, it must be repeated and reinforced regularly. For professional training it takes the dedication of the academic advisor, or refreshing the experience in other courses, to provide such reinforcement. Second, it takes an effort to get graduate students and their advisors to buy into professional education because the time needed for such education at the outset appears to them to decrease the time available for research. Our impression is that professional education makes students more efficient in their studies and research, even taking into account the time needed for the professional training. In the absence of hard data to substantiate this claim, however, it can be difficult to convince others of the value in devoting time for such professional training. This is aggravated by the fact that not all scientists appreciate the relevance of teaching and learning topics beyond disciplinary skills in their narrowest sense. Third, it does take resources, in particular time, to offer a broad professional education. Realistically, the required time and other resources are made available only when the institution acknowledges the importance of professional education.

Professional graduate training offers a number of advantages for students. First, being better prepared for scientific work can help minimize unnecessary frustration and loss of time both in graduate school and beyond. Second, it can help increase both the quality and quantity of the scientific work done in graduate school. Third, students learn skills to become better collaborators and to work more effectively with advisors. Fourth, such training should help students communicate their research more effectively. Fifth, broad professional training helps students be better prepared for the job market, and, sixth, it helps students to be better scientists.

Advantages also accrue to academic departments and their faculty members who offer professional training. Such training helps students to be more effective in their work, thus reducing the workload of advisors. For example, we discovered in our research group (the Center for Wave Phenomena) that offering students a class on academic publishing in combination with tutoring to improve their writing skills saved a large amount of time for advisors assisting students to write their thesis and publications. Moreover, offering an attractive program of professional education can help attract better students. As is generally appreciated, the quality of students is essential for the well-being of an academic research group; raising the level of incoming students elevates the scientific creativity and productivity of the whole group. Last the requirement of funding agencies to offer professional training is growing rapidly. The recent report of NIH (2011) points in this direction. Both NIH and NSF now require training in research ethics. The ‘broader impact’ criterion of the National Science Foundation is becoming increasingly important, and large programs such as the Integrative Graduate Education and Research Traineeship (IGERT) of NSF require professional education. Having an institutional program for professional education in place not only helps to offer such training, it obviates the need for individual faculty members to develop such training, and it increases the chances of success in funding of proposals.

More than any other reasons for offering professional education to graduate students, we owe it to our students to give them the best preparation possible to be the tomorrow’s professionals. Graduate students ought not be viewed as cheap labor to help us in our research; rather, the primary purpose of a graduate program is to educate young researchers who carry the torch of science forward and assure the continuity of the scientific endeavor in the best possible way. We should prepare them for this work as well as we can.

REFERENCES

Bronowski, J., 1956, Science and human values: Julian Messner Inc.

Covey, S., 1990, The 7 habits of highly effective people: Fireside Books.


Ivie, R., 2011, Got skills? On-the job activities for physicists: Presented at the Annual Meeting of the American Physical Society, Dallas TX.


Mervis, J., 2011, NIH report urges greater emphasis on training for all graduate students: Science, 331, 525.


