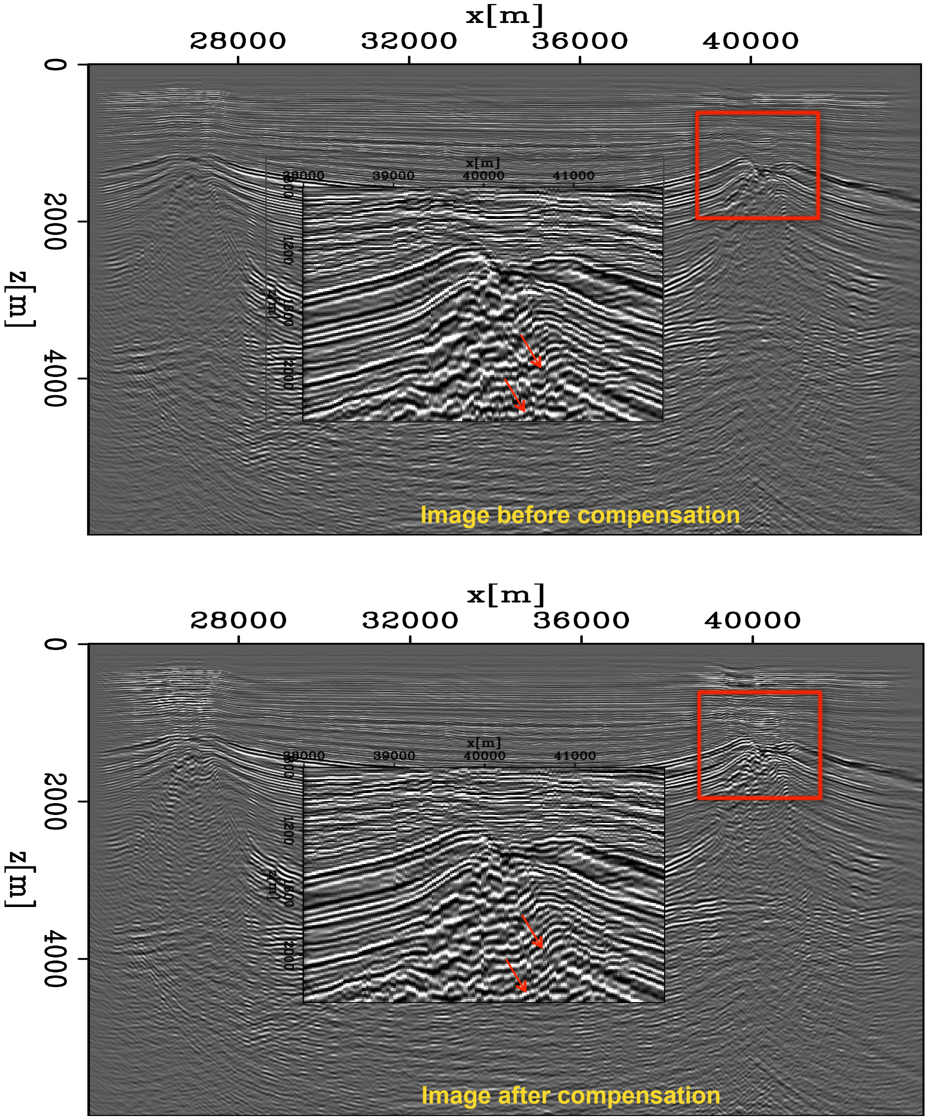
STANFORD EXPLORATION PROJECT

Ohad Barak, Biondo Biondi, Alejandro Cabrales-Vargas, Robert Clapp, Stuart Farris,
Fantine Huot, Joseph Jennings, Huy Le, Eileen Martin, Shuki Ronen, Rahul Sarkar,
Yi Shen, John Washbourne, and Ethan Williams

Report Number 165, October 2016



Preface

The electronic version of this report[1] makes the included programs and applications available to the reader. The markings [ER], [CR], and [NR] are promises by the author about the reproducibility of each figure result. Reproducibility is a way of organizing computational research that allows both the author and the reader of a publication to verify the reported results. Reproducibility facilitates the transfer of knowledge within SEP and between SEP and its sponsors.

ER denotes Easily Reproducible and are the results of processing described in the paper. The author claims that you can reproduce such a figure from the programs, parameters, and makefiles included in the electronic document. The data must either be included in the electronic distribution, be easily available to all researchers (e.g., SEG-EAGE data sets), or be available in the SEP data library[2]. We assume you have a UNIX workstation with Fortran, Fortran90, C, C++, X-Windows system and the software downloadable from our website (SEP makerules, SEPlib, and the SEP latex package), or other free software such as SU. Before the publication of the electronic document, someone other than the author tests the author's claim by destroying and rebuilding all ER figures. Some ER figures may not be reproducible by outsiders because they depend on data sets that are too large to distribute, or data that we do not have permission to redistribute but are in the SEP data library.

CR denotes Conditional Reproducibility. The author certifies that the commands are in place to reproduce the figure if certain resources are available. The primary reasons for the CR designation is that the processing requires 20 minutes or more, MPI or CUDA based code, or commercial packages such as Matlab or Mathematica.

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

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

---

[1] http://sepwww.stanford.edu/private/docs/sep165

[2] http://sepwww.stanford.edu/public/docs/sepdatalib/toc_html

[3] http://sepwww.stanford.edu/research/redoc/

SEP165 — TABLE OF CONTENTS

Anisotropy and attenuation

Inversion

Machine learning and event detection

Software

Stanford Exploration Project, SEP165, October 8, 2016

Practical issues in anisotropic full waveform inversion

Huy Le

ABSTRACT

In the previous report, SEP 163, I demonstrated the potential of using the second-order pseudo-acoustic anisotropic wave equations for full waveform inversion with simple examples. In this report, I investigate some practical issues that anisotropic full waveform inversion implemented with this system of equations faces. These issues are parameter sensitivity, parameterization, and null space. Firstly, to take into account differences in sensitivity of different anisotropic parameters, I suggest a simple normalization technique. This normalization results in dimensionless parameters that better reflect sensivity to recorded data. Secondly, a number of parameterizations, including stiffness coefficients, velocities, and Thomsen's parameters, are tested on the BP 2007 synthetic model. I found that parameterizations with one velocity and two Thomsen's parameters gave the best results in terms of reduction in the objective function. Lastly, to mitigate the null space problem, I regularize the inversion with steering filters based on the migrated image's dip information.

INTRODUCTION

Full waveform inversion (FWI) is a powerful technique to obtain both the long-wavelength (background) and the short-wavelength (reflectivity) components of a velocity model by minimizing the difference between the observed data and the modeled data. Successful applications of FWI generally require low-frequency and long-offset data, particularly diving waves, in order to solve the cycle skipping problem. When processing long-offset data, it is necessary to incorporate anisotropy to account for the dependence of seismic velocity on propagation direction. Studies have shown that FWI with anisotropy leads to better results (Barnes et al., 2008; Lee et al., 2010).

Anisotropic FWI poses more challenges compared to isotropic FWI. These challenges come from the increase in the number of unknown parameters. An anisotropic medium is characterized not only by velocity but also by anisotropic parameters that describe how velocity changes with direction. The first challenge is to account for different sensitivities, units, and magnitudes among these different parameters. The second challenge is to find a suitable parameterization. Multiple parameters are required for anisotropy and there are a number of equivalent sets of them. Another challenge in anisotropic FWI is the extension of the problem's null space, which is a direct consequence of increasing number of parameters. In this report, I propose some solutions to these challenges.

Because the earth supports elastic wave motions, one will have to use elastic wave equations in order to handle the amplitude and kinematics of seismic waves. However, in the context of parameter estimation, a large number of unknowns in elastic FWI poses many difficulties, both theoretically and computationally. Consequently, acoustic wave equations have been commonly used and have shown successes in many FWI applications (Gholami et al., 2013b; Warner et al., 2013). Although acoustic anisotropy is not realistic, the resulting equations may accurately describe the kinematics of waves in anisotropic media. This work adopts a time-domain method that was developed by Le (2016) in the previous report for anisotropic FWI. This method is based on the second-order pseudo-acoustic wave equations in transverse isotropic media (VTI) (Duveneck and Bakker, 2011):

$$\begin{cases} \partial_t^2 p = c_{11}\partial_x^2 p + c_{13}\partial_z^2 q + f_x, \\ \partial_t^2 q = c_{13}\partial_x^2 p + c_{33}\partial_z^2 q + f_z. \end{cases} \tag{1}$$

p and q are the normal stresses in the x-direction and z-direction, f_i are the sources, and c_{ij} are the density-normalized stiffness coefficients.

Define $\mathbf{x}$ to be the location vector, $\mathbf{x}_r$ to be the receiver location, and $\mathbf{m}$ to be the vector of model parameters. The modeled pressure field is defined as the average stress, $d(\mathbf{x}, t; \mathbf{m}) = \frac{1}{2}(p + q)$. The FWI objective function can be defined as the l_2-norm of the difference between the modeled data and the observed data, $d_0(\mathbf{x}_r, t)$:

$$\chi(\mathbf{m}) = \frac{1}{2}\|d(\mathbf{x}_r, t; \mathbf{m}) - d_0(\mathbf{x}_r, t)\|_2^2. \tag{2}$$

Following the adjoint state method presented in Fichtner (2011), the gradients of the objective function with respect to model parameters are computed as cross-correlations of the adjoint wavefields, p_1 and q_1, and the forward modeling wavefields, p and q:

$$g_{c_{11}} = \int_0^T p_1 \partial_x^2 p dt, \tag{3a}$$

$$g_{c_{13}} = \int_0^T (p_1 \partial_z^2 q + q_1 \partial_x^2 p) dt, \tag{3b}$$

$$g_{c_{33}} = \int_0^T q_1 \partial_z^2 q dt. \tag{3c}$$

The adjoint wavefields are solutions to the adjoint equations:

$$\begin{cases} \partial_t^2 p_1 = \partial_x^2 (c_{11} p_1 + c_{13} q_1) + \frac{1}{2}(d - d_0)\delta(\mathbf{x} - \mathbf{x_r}), \\ \partial_t^2 q_1 = \partial_z^2 (c_{13} p_1 + c_{33} q_1) + \frac{1}{2}(d - d_0)\delta(\mathbf{x} - \mathbf{x_r}). \end{cases} \tag{4}$$

BALANCING THE GRADIENTS

There are several ways to parameterize the subsurface and the inversion problem. One common way is to use velocity and Thomsen's anisotropic parameters (Thomsen, 1986). The stiffness coefficients, c_{ij}, are related to vertical P-velocity, v_z, ϵ, and δ, by:

$$c_{11} = v_z^2(1 + 2\epsilon), \tag{5a}$$

$$c_{13} = v_z^2\sqrt{1 + 2\delta}, \tag{5b}$$

$$c_{33} = v_z^2. \tag{5c}$$

The gradients of the FWI objective function with respect to v_z, ϵ, and δ, can be easily computed from equations 3 by chain rules:

$$g_{v_z} = 2v_z(1+2\epsilon)g_{c_{11}} + 2v_z\sqrt{1+2\delta}g_{c_{13}} + 2v_z g_{c_{33}}, \tag{6a}$$

$$g_{\epsilon} = 2v_z^2(1+2\epsilon)g_{c_{11}}, \tag{6b}$$

$$g_{\delta} = \frac{v_z^2}{\sqrt{1+2\delta}}g_{c_{13}}. \tag{6c}$$

Because the stiffness coefficients, c_{ij}, have the same dimension, the same order of magnitude, and approximately equal sensitivity to the observed data (Gholami et al., 2013a), equations 3 can be used directly in a gradient-based optimization algorithm to minimize the objective function (Le, 2016). When parameterized with v_z, ϵ, and δ, these parameters not only have different dimensions and magnitudes, but also very different sensitivities, with v_z the most sensitive parameter. Its sensitivity is an order of magnitude larger than that of ϵ, which is in turn much larger than δ's. As a result, not accounting for these differences can lead to bad estimates of subsurface parameters.

I propose a simple normalization that can mitigate the differences in dimensions, magnitudes, and sensitivities among parameters such as v_z, ϵ, and δ. Instead of inverting for these parameters directly, I invert for a set of equivalent but normalized parameters, $\bar{v}_z$, $\bar{\epsilon}$, and $\bar{\delta}$ defined by:

$$\bar{v}_z = \frac{v_z}{v_0}, \bar{\epsilon} = \frac{\epsilon}{\epsilon_0}, \bar{\delta} = \frac{\delta}{\delta_0}. \tag{7}$$

The normalization coefficients, v_0, ϵ_0, and δ_0 are scalar constants chosen to satisfy:

$$\frac{g_{\bar{v}_z}}{\bar{v}_z} = \frac{g_{\bar{\epsilon}}}{\bar{\epsilon}} = \frac{g_{\bar{\delta}}}{\bar{\delta}}. \tag{8}$$

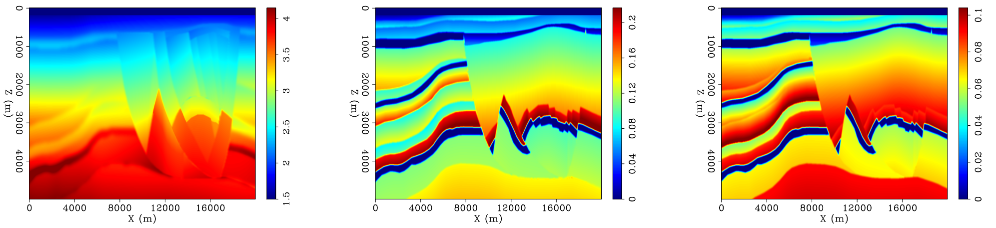
These equations promote approximately the same amount of update to all parameters at each inversion iteration.

I illustrate the improvement of this parameter normalization using a portion of the BP 2007 VTI synthetic model. Figure 1 shows the true model parameters and Figure 2 shows the initial model parameters. The observed data is synthesized using the same engine as the modeled data with a Ricker wavelet of 5 Hz central frequency. 100 sources and 1600 receivers are placed uniformly on the water surface. Source spacing is 200 meters and receiver spacing is 12.5 meters. Maximum offset is 20 km.

Figure 3 shows the gradients with respect to unnormalized parameters, v_z, ϵ, and δ. Note from the scale bars that g_{v_z} (Figure 3(a)) is three orders of magnitude smaller than the other two gradients, g_{ϵ} (Figure 3(b)) and g_{δ} (Figure 3(c)). After normalization, $g_{\bar{v}_z}$ (Figure 4(a)) is now greater than $g_{\bar{\epsilon}}$ (Figure 4(b)), which is greater than $g_{\bar{\delta}}$ (Figure 4(c)). This means that the gradients with respect to normalized parameters better reflect their sensitivities to the data. Notice that because the normalization coefficients v_0, ϵ_0, and δ_0 are scalar constants, the change of variables from unnormalized to normalized does not change the geologic structures in the gradients, but only the magnitudes of their updates.

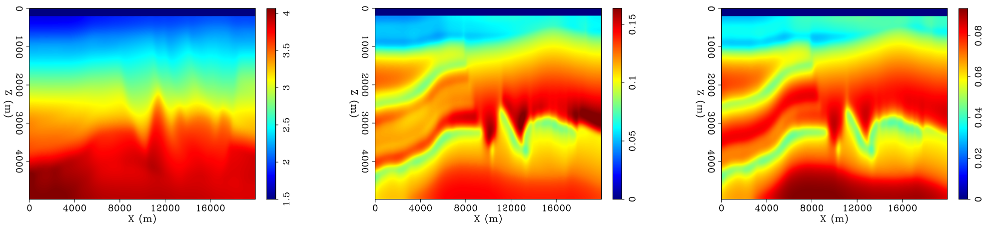
I performed two inversions with unnormalized and normalized parameters, and the results are shown in Figures 5 and 6 respectively. Compared with the true model (Figure 1), the inverted model using unnormalized parameters (Figure 5) shows unrealistic updates in

ϵ and δ and almost no update in v_z. Inversion with normalized parameters (Figure 6) shows significant update in velocity, although updates in ϵ and δ are small. Figure 7 compares two objective functions from these inversions. Inversion with normalized parameters reduces the objective function much more than without normalization. In fact, without normalization, the inversion gets stuck after 13 iterations.



(a) (b) (c)

Figure 1: True model: (a) v_z in km/s, (b) ϵ, and (c) δ. [**ER**] huyle/. v.bp2007,eps.bp2007,del.bp2007



(a) (b) (c)

Figure 2: Initial model: (a) v_z in km/s, (b) ϵ, and (c) δ. [**ER**] huyle/. bgv.bp2007,bgeps.bp2007,bgdel.bp2007

PARAMETERIZATION TEST

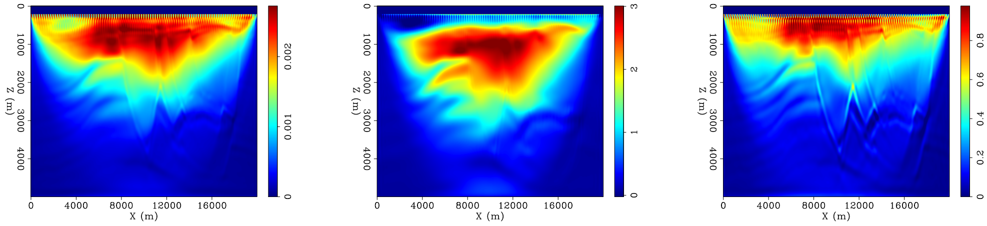
Parameterization is a vital aspect of anisotropic FWI. A set of parameters that is most resolvable and has minimal trade-offs can lead to better inversion results. Besides stiffness coefficients c_{ij}, vertical velocity v_z, and Thomsen's parameters ϵ and δ, there are other parameters that also play an important role in describing anisotropic media. They are horizontal velocity v_h, normal moveout (NMO) velocity v_n, and the anellipticity parameter η:

$$v_h = v_z\sqrt{1+2\epsilon}, \tag{9a}$$

$$v_n = v_z\sqrt{1+2\delta}, \tag{9b}$$

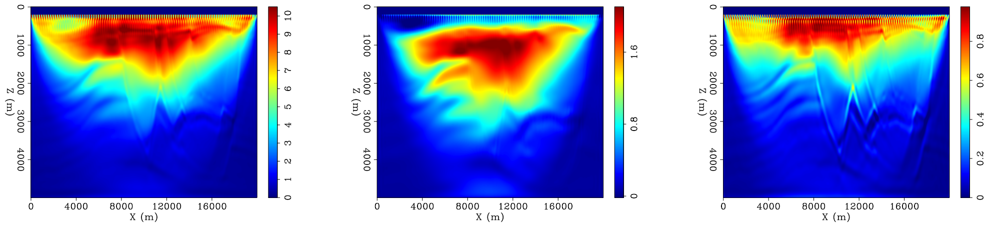
$$\eta = \frac{\epsilon-\delta}{1+2\delta}. \tag{9c}$$

Many studies have been devoted to parameterization in anisotropic media. Alkhalifah and Tsvankin (1995) derive an expression for P-wave travel time in VTI media and show that it depends on two parameters: NMO velocity v_n and anellipticity parameter η. Plessix



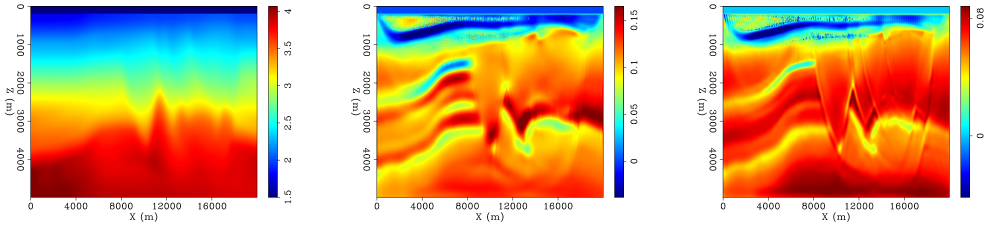
(a) (b) (c)

Figure 3: Gradients with respect to unnormalized parameters: (a) g_{v_z}, (b) g_ϵ, and (c) g_δ. Notice from the scale bars that g_{v_z} is three orders of magnitude smaller than the other two gradients. [**CR**] huyle/. gv0,geps0,gdel0



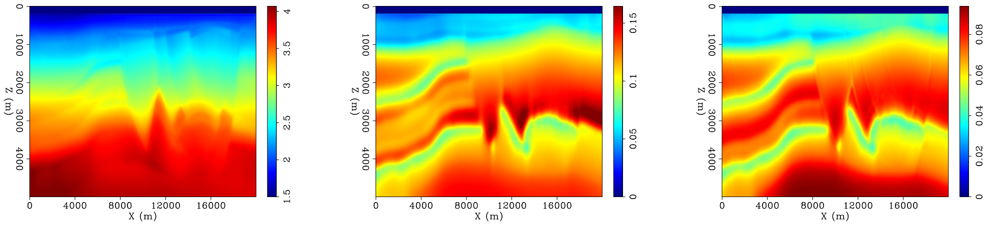
(a) (b) (c)

Figure 4: Gradients with respect to normalized parameters: (a) $g_{\bar{v}_z}$, (b) $g_{\bar{\epsilon}}$, and (c) $g_{\bar{\delta}}$. The scale bars show that these gradients more accurately reflect the sensitivities of unknown parameters. [**CR**] huyle/. gv,geps,gdel



(a) (b) (c)

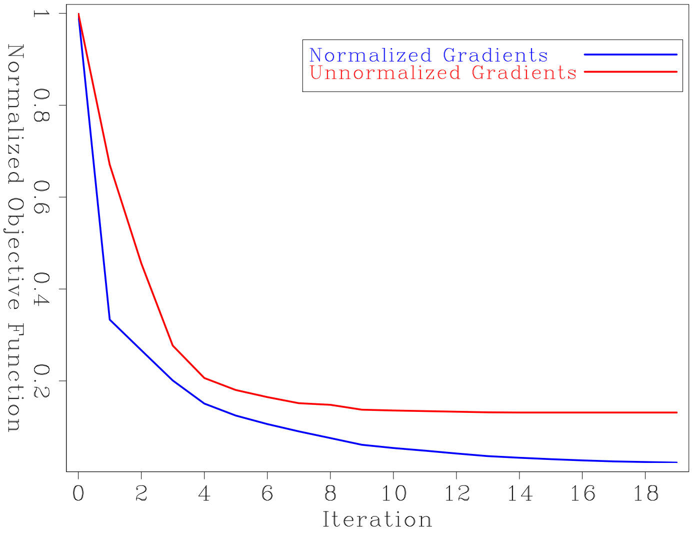
Figure 5: Inverted model using unnormalized parameters: (a) v_z in km/s, (b) ϵ, and (c) δ. Velocity is almost not updated while the other two parameters show unrealistic features. [**CR**] huyle/. v0.1,eps0.1,del0.1



(a) (b) (c)

Figure 6: Inverted model using normalized parameters: (a) v_z in km/s, (b) ϵ, and (c) δ. Velocity is much better resolved compared to the inversion with unnormalized parameters, even though anisotropic parameters are weakly updated. [**CR**] huyle/. v.1,eps.1,del.1

Figure 7: Comparison of objective functions from inversions with and without normalization. Inversion with normalized parameters does not get stuck and leads to a lower objective function. [**CR**] huyle/. objfuncs1



and Cao (2011) compute the eigenvalues of the Hessian of the FWI objective function and conclude that NMO and horizontal velocities are the most sensitive parameters. While NMO velocity is responsible for reflection events, horizontal velocity is responsible for diving waves and post-critical reflections. Gholami et al. (2013a) analyze and plot the radiation patterns of different sets of parameters to study sensitivity and trade-off.

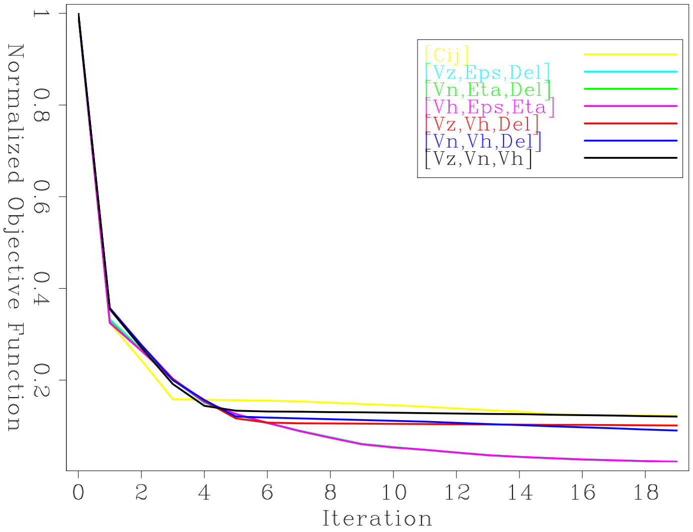
In this work, I carry out several inversions using parameterizations that have been studied by Gholami et al. (2013a) in order to compare their performances. Four types of parameterizations are tested. The first type includes stiffness coefficients c_{ij}. The second type consists of three parameterizations with one velocity and two anisotropic parameters: $\{v_z, \epsilon, \delta\}$, $\{v_n, \eta, \delta\}$, and $\{v_h, \epsilon, \eta\}$. The third type has two velocities and one anisotropic parameter in two different sets: $\{v_z, v_h, \delta\}$ and $\{v_n, v_h, \delta\}$. The last type contains three velocities $\{v_z, v_n, v_h\}$. The inversions are run on the BP 2007 synthetic model with the same setting and acquisition as mentioned earlier. They start from the same initial model and the normalization technique described in the previous section is applied whenever unknown parameters are of different dimensions (the second and third types).

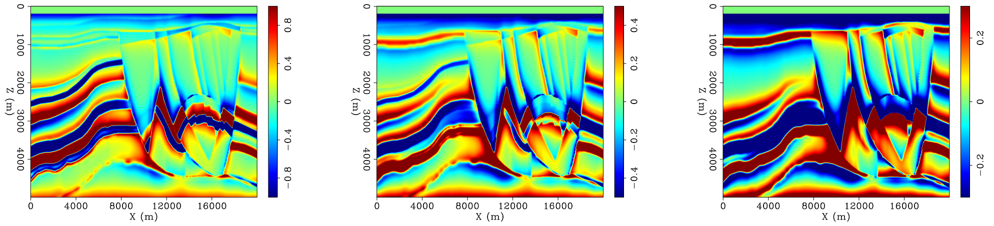
Figure 8 shows a comparison of objective functions from inversions with different parameterizations. I observe that the second type of parameterization with one velocity and two anisotropic parameters gave the best results in terms of reduction in objective function. Within this type, all three parameterizations, $\{v_z, \epsilon, \delta\}$, $\{v_n, \eta, \delta\}$, and $\{v_h, \epsilon, \eta\}$, perform equally well. The third type of parameterization with two velocities and one anisotropic parameter is second in performance while parameterizations with stiffness coefficients (the first type) and with three velocities (the fourth type) give roughly equally worst results.

Figures 10, 12, 14, 16, 18, 20, and 22 show updates of model parameters for seven tested parameterizations. These updates are to be compared with the corresponding differences between the true and the initial model parameters, which are shown in Figures 9, 11, 13, 15, 17, 19, and 21 respectively. Some observations can be made. Firstly, parameterizations with one velocity and two anisotropic parameters always give the best velocity estimates. NMO velocity, v_n, is best inverted using parameterization $\{v_n, \eta, \delta\}$ (Figure 14(a)), while horizontal velocity, v_h, is best resolved using $\{v_h, \epsilon, \eta\}$ (Figure 16(a)), and vertical velocity, v_z, is best resolved using parameterization $\{v_z, \epsilon, \delta\}$ (Figure 12(a)). However, the inverted velocity models show some unreal features that is a result of cross-talk from two anisotropic parameters within their parameterizations.

I also observe that parameterization $\{v_z, \epsilon, \delta\}$ also gives the best estimates in ϵ and δ (Figures 12(b) and 12(c) respectively). However, update in δ is very weak and contaminated by cross-talk from velocity. In fact, this is true for all parameterizations that include δ. Particularly, this cross-talk from velocity is so strong that δ is updated in the wrong direction in parameterizations $\{v_n, \eta, \delta\}$ (Figures 14(c)) and $\{v_n, v_h, \delta\}$ (Figure 20(c)). In terms of η, it is best resolved using parameterization $\{v_n, \eta, \delta\}$ (Figure 14(b)). However, the inverted η parameter is again degraded by cross-talk from velocity (NMO velocity, v_n, in this case), especially in the shallow part of the model. The inter-parameter cross-talk between velocity and anisotropic parameters is expected because the wide range of scattering angle over which velocity is resolvable overlap with those angles of anisotropic parameters (Gholami et al., 2013a).

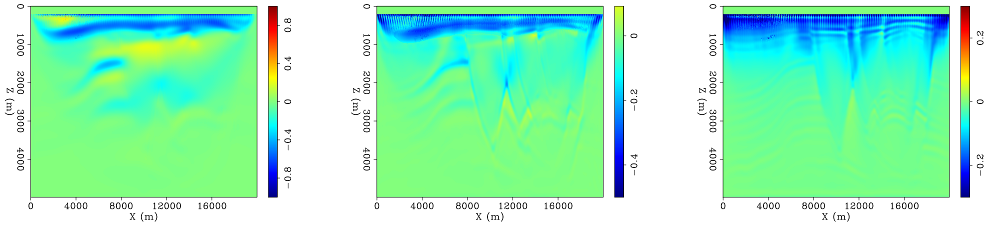
Figure 8: Comparison of objective functions from inversions with different parameterizations. Parameterizations with one velocity and two anisotropic parameters lead to the most reduced objective functions. [**CR**] huyle/. objfuncs2





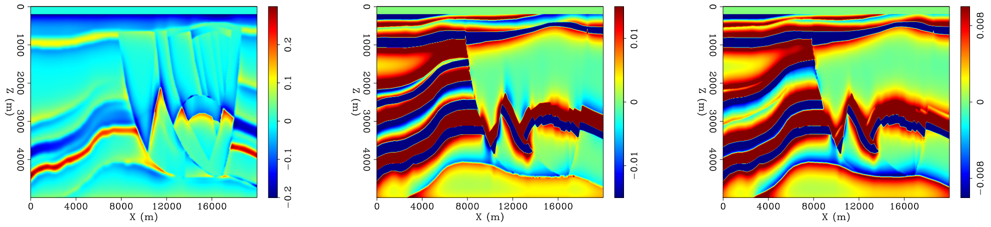
(a) (b) (c)

Figure 9: Differences between initial and true model: (a) c_{11}, (b) c_{13}, and (c) c_{33} in MPa. [**ER**] huyle/. dc11.bp2007,dc13.bp2007,dc33.bp2007



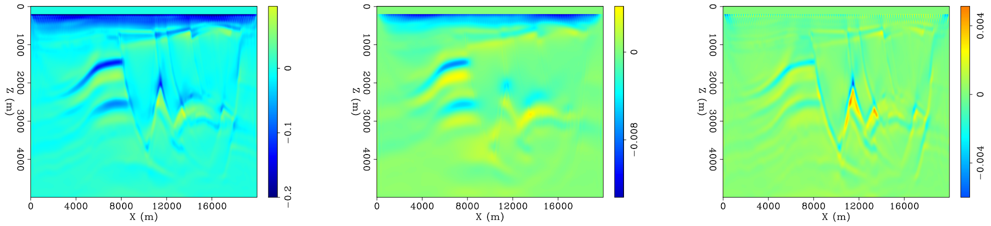
(a) (b) (c)

Figure 10: Updates after inversion: (a) c_{11}, (b) c_{13}, and (c) c_{33} in MPa. These updates are to be compared with the corresponding differences between true and initial model in Figure 9. [**CR**] huyle/. dc11,dc13,dc33



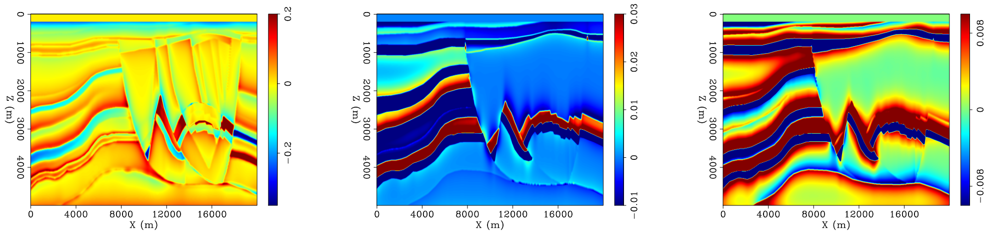
(a) (b) (c)

Figure 11: Differences between initial and true model: (a) v_z in km/s, (b) ϵ, and (c) δ. [**ER**] huyle/. dv.bp2007,deps.bp2007,ddel.bp2007



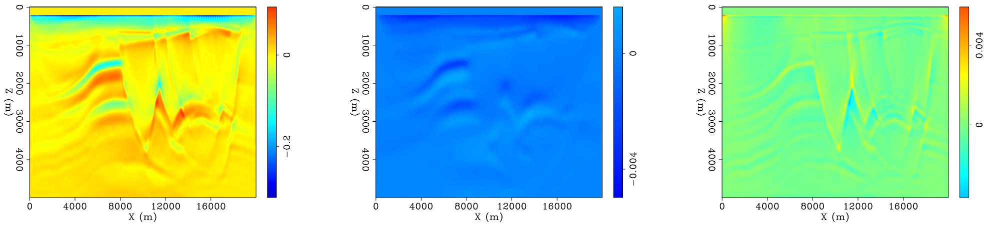
(a) (b) (c)

Figure 12: Updates after inversion: (a) v_z in km/s, (b) ϵ, and (c) δ. These updates are to be compared with the corresponding differences between true and initial model in Figure 11. This parameterization gives the best estimates in both velocity and anisotropic parameters. There is some cross-talk between velocity and the other two parameters, especially δ. [**CR**] huyle/. dv.1,deps.1,ddel.1



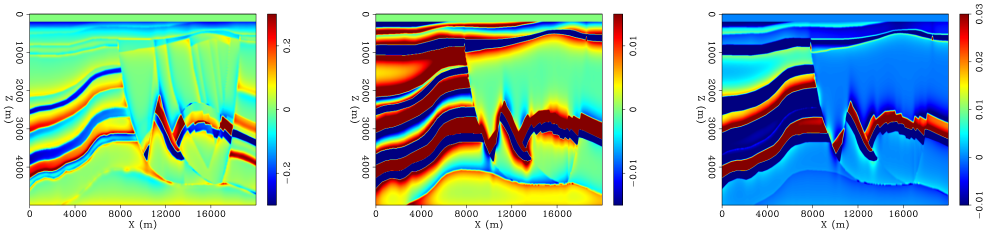
(a) (b) (c)

Figure 13: Differences between initial and true model: (a) v_n in km/s, (b) η, and (c) δ. [**ER**] huyle/. dvn.bp2007,deta.bp2007,ddel.bp2007



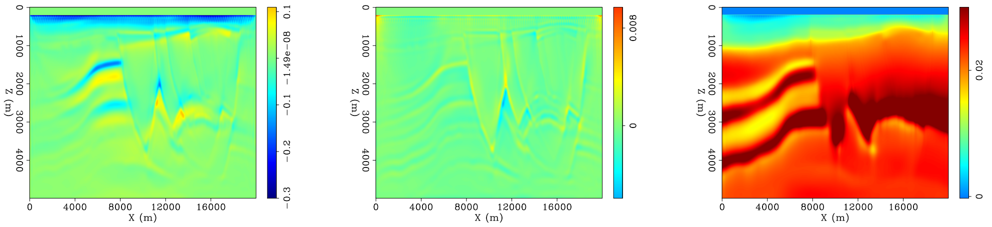
(a) (b) (c)

Figure 14: Updates after inversion: (a) v_n in km/s, (b) η, and (c) δ. These updates are to be compared with the corresponding differences between true and initial model in Figure 13. This parameterization gives the best estimates in v_n and η. Inverted models of theses two parameters also show some cross-talk. Cross-talk from v_n leads to wrong-direction update in δ. [**CR**] huyle/. dvn.2,deta.2,ddel.2



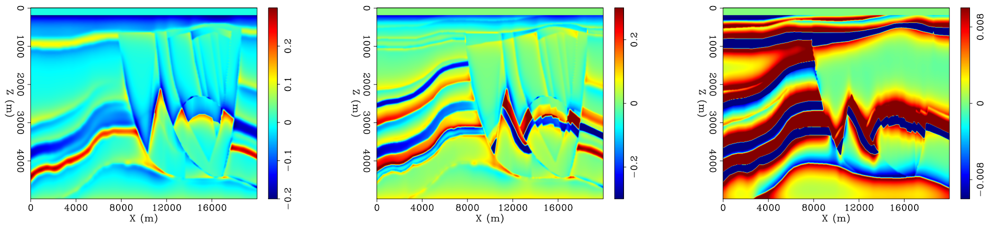
(a) (b) (c)

Figure 15: Differences between initial and true model: (a) v_h in km/s, (b) ϵ, and (c) η. [**ER**] huyle/. dvh.bp2007,deps.bp2007,deta.bp2007



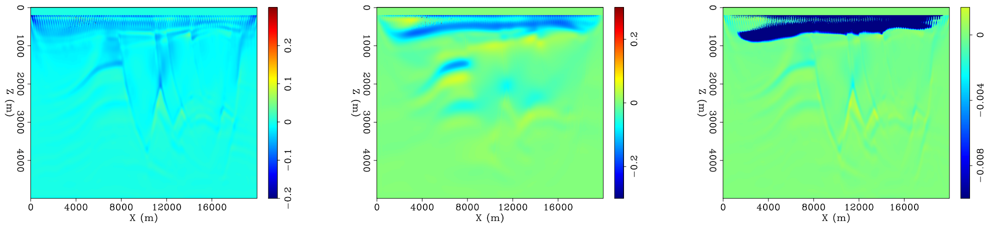
(a) (b) (c)

Figure 16: Updates after inversion: (a) v_h in km/s, (b) ϵ, and (c) η. These updates are to be compared with the corresponding differences between true and initial model in Figure 15. This parameterization gives the best estimate in v_h. ϵ update is contaminated by v_h. η is updated in the wrong direction. [**CR**] huyle/. dvh.3,deps.3,deta.3



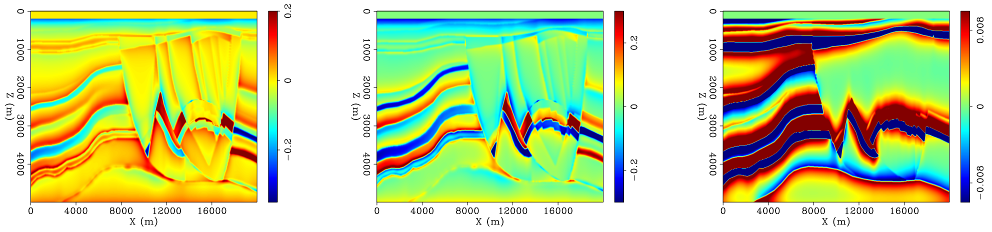
(a) (b) (c)

Figure 17: Differences between initial and true model: (a) v_z in km/s, (b) v_h in km/s, and (c) δ. [**ER**] huyle/. dv.bp2007,dvh.bp2007,ddel.bp2007



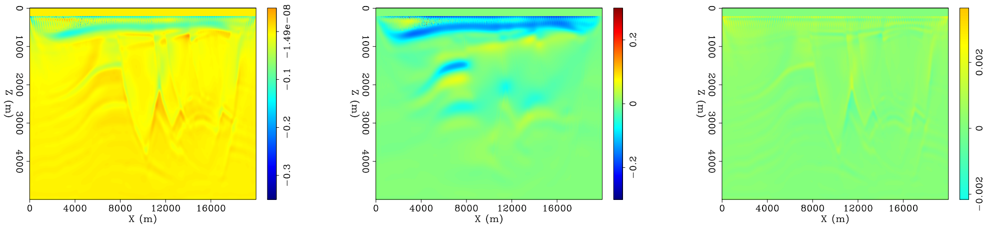
(a) (b) (c)

Figure 18: Updates after inversion: (a) v_z in km/s, (b) v_h in km/s, and (c) δ. These updates are to be compared with the corresponding differences between true and initial model in Figure 17. [**CR**] huyle/. dv.4,dvh.4,ddel.4



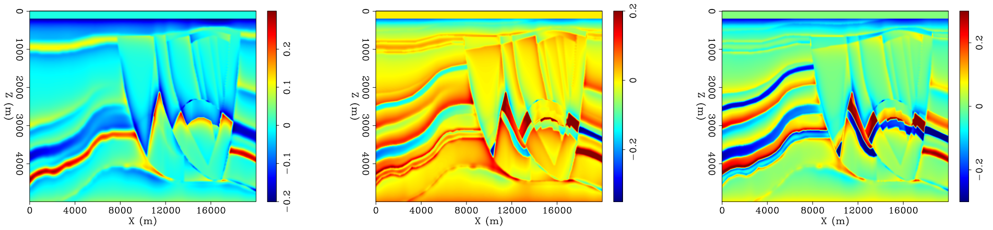
(a) (b) (c)

Figure 19: Differences between initial and true model: (a) v_n in km/s, (b) v_h in km/s, and (c) δ. [**ER**] huyle/. dvn.bp2007,dvh.bp2007,ddel.bp2007



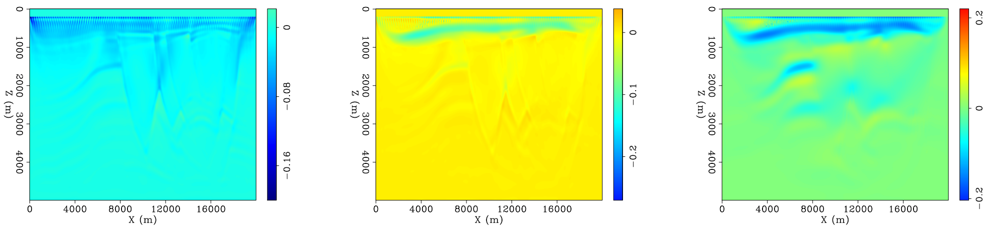
(a) (b) (c)

Figure 20: Updates after inversion: (a) v_n in km/s, (b) v_h in km/s, and (c) δ. These updates are to be compared with the corresponding differences between true and initial model in Figure 19. In some areas, δ is updated in the wrong direction due to cross-talk from velocities. [**CR**] huyle/. dvn.5,dvh.5,ddel.5



(a) (b) (c)

Figure 21: Differences between initial and true model: (a) v_z, (b) v_n, and (c) v_h in km/s. [**ER**] huyle/. dv.bp2007,dvn.bp2007,dvh.bp2007



(a) (b) (c)

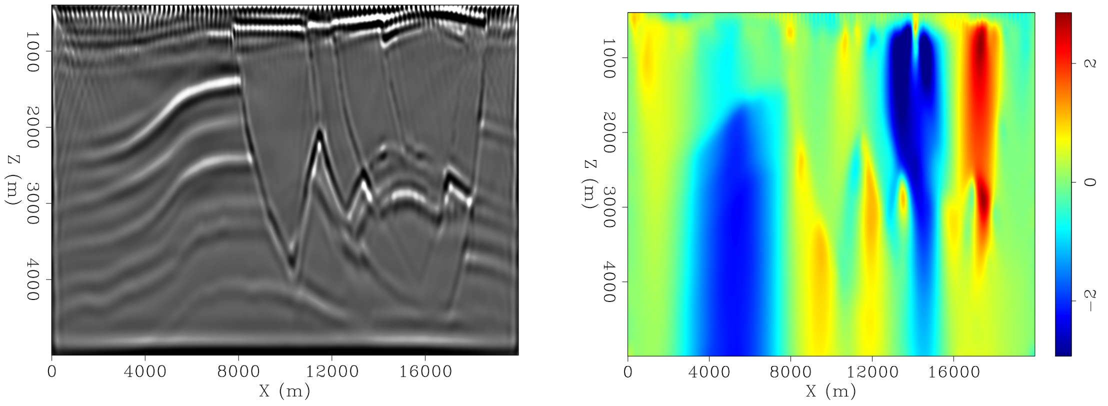
Figure 22: Updates after inversion: (a) v_z, (b) v_n, and (c) v_h in km/s. These updates are to be compared with the corresponding differences between true and initial model in Figure 21. [**CR**] huyle/. dv.6,dvn.6,dvh.6

REGULARIZATION

Although acoustic anisotropic FWI has a smaller number of model parameters than its elastic equivalent, compared to isotropic FWI, its model space is still much larger. This results in a much bigger model null space, making the problem more ill-posed, and the inversion can more easily get stuck in local minima. One standard way to overcome this issue is to use regularization. The regularized objective function is expressed as:

$$\chi(\mathbf{m}) = \frac{1}{2}\|d(\mathbf{x}_r, t; \mathbf{m}) - d_0(\mathbf{x}_r, t)\|_2^2 + \frac{\alpha}{2}\|\mathbf{A}\mathbf{m}\|_2^2, \tag{10}$$

in which $\mathbf{A}$ is some styling operator that is related to the model covariance matrix. In the next example, I take this operator to be a steering filter obtained from the dip of the migrated image. Using a steering filter to spread dip information along geologic structures helps speed up convergence and improve inversion results. This was introduced and applied successfully to tomography by Clapp et al. (2004). Figure 23(a) shows the migrated image and Figure 23(b) shows the tangent of the dip angle, which is used to construct the steering filter $\mathbf{A}$. The inversion results with regularization are shown in Figure 24. This figure shows improvements in the inverted models compared to the non-regularized results (Figure 12). Dipping reflectors are more continuous and the inverted model parameters are closer to the true ones. Figure 25 shows that regularized inversion leads to a slightly lower objective function than without regularization. The improvements are, however, not very significant, which I suspect is due to the smooth dip field (Figure 23(b)). Moreover, cross-talk between velocity and anisotropic parameters is still present because the same steering filter is applied on all three parameters. Additional information (from rock physics, for example) can be used to obtain separate filters for individual parameters that can reduce their cross-talk.



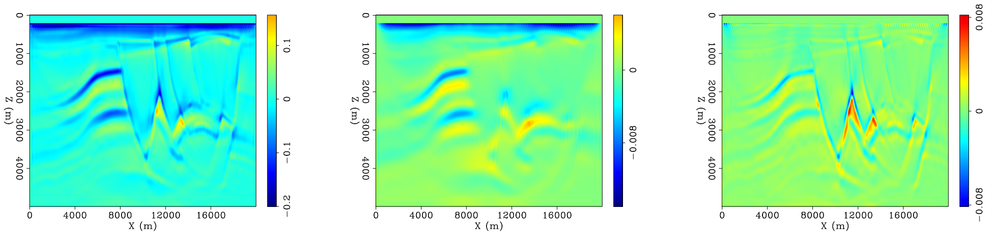
(a)

(b)

Figure 23: (a) Migrated image and (b) dip (tangent of dip angle). [**CR**] huyle/. image,dip

CONCLUSIONS

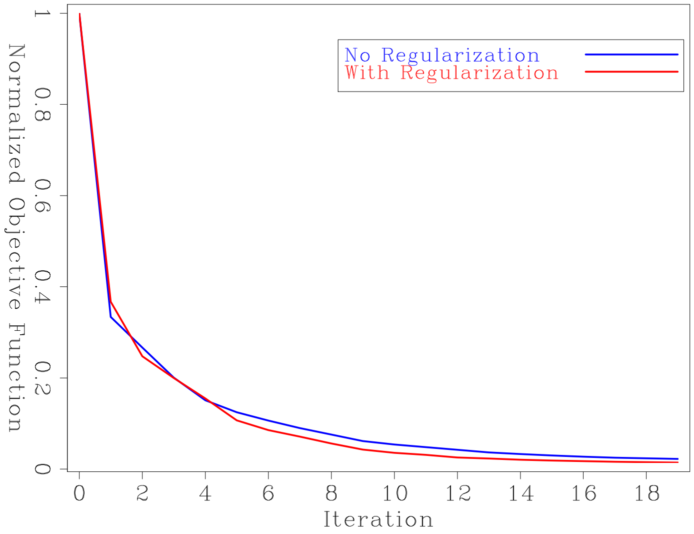
In this report I investigate three practical issues in anisotropic FWI. The first issue is how to account for differences in dimension, magnitude, and sensitivity among different parameters. I suggest applying a normalization which transforms the original parameters into a



(a) (b) (c)

Figure 24: Updates after inversion with steering filter regularization: (a) v_z in km/s, (b) ϵ, and (c) δ. Inverted model parameters show more continuous dipping structures and better converge toward the true ones. Cross-talk is, however, still observable. [**CR**] huyle/. dv.1r,deps.1r,ddel.1r

Figure 25: Comparison of objective functions from inversions with and without regularization. Regularization leads to a slightly lower objective function. [**CR**] huyle/. objfuncs3



set of dimensionless parameters that more truly reflect their sensitivities. This normalization can be seen as a simple way to approximate the Hessian matrix. The second issue is parameterization. I have tested four types of parameterization: stiffness coefficients, one velocity with two anisotropic parameters, two velocities with one anisotropic parameters, and three velocities. Performance of different parameterizations is evaluated by how much they reduce the objective function. I found that parameterizations with one velocity and two anisotropic parameters gave the best results. Especially, among all tested parameterizations, $\{v_z, \epsilon, \delta\}$ results in the best estimates in both vertical velocity and anisotropic parameters despite some cross-talk. Another issue in anisotropic FWI is the large model null space. I have shown that a regularization with a steering filter constructed using dip information from the migrated image can improve the results and lead to better inverted models. Additional information from rock physics can also be used for regularization and is the subject of my future work.

REFERENCES

Alkhalifah, T. and I. Tsvankin, 1995, Velocity analysis for transversely isotropic media: Geophysics, **60**, 1550–1566.

Barnes, C., M. Charara, and T. Tsuchiya, 2008, Feasibility study for an anisotropic full waveform inversion of crosswell seismic data: Geophysical Prospecting, **56**, 897–906.

Clapp, R. G., B. Biondi, and J. F. Claerbout, 2004, Incorporating geologic information into reflection tomography: Geophysics, **69**, 533–546.

Duveneck, E. and P. M. Bakker, 2011, Stable P-wave modeling for reverse-time migration in tilted TI media: Geophysics, **76**, No. 2, WA3–WA11.

Fichtner, A., 2011, Full seismic waveform modeling and inversion: Springer.

Gholami, Y., R. Brossier, S. Operto, A. Ribodetti, and J. Virieux, 2013a, Which parameterization is suitable for acoustic vertical transver isotropic full waveform inversion? Part 1: Sensitivity and trade-off analysis: Geophysics, **78**, No. 2, R81–R105.

———, 2013b, Which parameterization is suitable for acoustic vertical transver isotropic full waveform inversion? Part 2: Synthetic and real data case studies from Valhall: Geophysics, **78**, No. 2, R107–R124.

Le, H., 2016, Anisotropic full waveform inversion : SEP-Report, **163**, 155–162.

Lee, H.-Y., J. M. Koo, D.-J. Min, B.-D. Kwo, and H. S. Yoo, 2010, Frequency-domain elastic full waveform inversion for VTI media: Geophysical Journal International, **183**, 884–904.

Plessix, R. E. and Q. Cao, 2011, A parameterization study for surface seismic full waveform inversion in an acoustic vertical transversely isotropic medium: Geophysical Journal International, **185**, 539–556.

Thomsen, L., 21986, Weak elastic anisotropy: Geophysics, **51**, 1954–1966.

Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Stekl, L. Guasch, C. Win, G. Conroy, and A. Bertrand, 2013, Anisotropic 3D full-waveform inversion: Geophysics, **78**, No. 2, R59–R80.

Stanford Exploration Project, SEP165, October 8, 2016

Reverse-time migration using the energy imaging condition in TTI media

Alejandro Cabrales-Vargas

ABSTRACT

I extend my previous work about the energy imaging condition for anisotropic reverse-time migration to tilted-transverse isotropic media. Synthetic examples show the feasibility of this implementation.

INTRODUCTION

This report constitutes the follow-up of a previous report (Cabrales-Vargas, 2016) where I implemented the energy imaging condition (EIC) in vertical-transverse isotropic (VTI) media. Similar to its isotropic counterpart, EIC can be used to enhance either the tomographic component (applicable to improving the full waveform inversion gradient for cycle skipping reduction) or the reflectivity component of the image (Whitmore and Crawley, 2012; Rocha et al., 2016a,b) for imaging purposes.

In this report I first introduce the EIC for tilted-trasverse isotropic (TTI) media. Next, I show the numerical results obtained on the BP TTI 2-D synthetic model. Finally, I present the conclusions of this work.

METHOD

In isotropic acoustic media (e.g. Rocha et al., 2016a) the EIC can be derived from the energy function of the acoustic wave equation, which is defined as

$$E(t) = \int_\Omega \left[\frac{1}{v^2} \left(\frac{\partial u}{\partial t} \right)^2 + \|\nabla u\|^2 \right] d\mathbf{x} = \int_\Omega \|\Box u\|^2 d\mathbf{x}, \tag{1}$$

where $E(t)$ is the energy of the acoustic wave equation, u is the propagation wavefield. From this expression Rocha et al. (2016a) derived the EIC as

$$I(\mathbf{x}) = \sum_t \Box S(\mathbf{x}, t) \cdot \widetilde{\Box} R(\mathbf{x}, t), \tag{2}$$

where S and R are the source wavefield and the receiver wavefield, respectively. The operator $\Box$ represents a space-time gradient (also known as *D'Alembertian*), given in three-dimensional space-time by

$$\Box = \left(\frac{1}{v} \frac{\partial}{\partial t}, \ \frac{\partial}{\partial x}, \ \frac{\partial}{\partial z} \right), \tag{3}$$

and the operator $\widetilde{\Box}$ is a modified space-time gradient given by

$$\widetilde{\Box} = \left(\cos(2\gamma_c) \frac{1}{v} \frac{\partial}{\partial t}, \ \frac{\partial}{\partial x}, \ \frac{\partial}{\partial z} \right), \tag{4}$$

where γ_c represents a cut-off angle that rejects seismic events in the neighborhood of such value for the reflection angle. For example, setting $\gamma_c = 0^\circ$ preserves the reverse-time migration (RTM) tomographic component and rejects reflections in the neighborhood of zero degrees. On the contrary, setting $\gamma_c = 90^\circ$ rejects most of the tomographic component and preserves the reflectivity component, similar to the Laplacian low-cut filter (Rocha et al., 2016a).

For the TTI case I utilize the anisotropic wave equations proposed by Zhang et al. (2011),

$$\frac{1}{v^2}\frac{\partial^2 \mathbf{p}}{\partial t^2} = \mathbf{B}\mathbf{G}_2\mathbf{p}, \tag{5}$$

where $\mathbf{p} = [p, r]^T$ is the stress vector with horizontal component, p, and vertical component, r, matrices $\mathbf{B}$ and $\mathbf{G}_2$ are given by

$$\mathbf{B} = \begin{bmatrix} 1+2\epsilon & \sqrt{1+2\delta} \\ \sqrt{1+2\delta} & 1 \end{bmatrix} \tag{6}$$

and

$$\mathbf{G}_2 = \begin{bmatrix} G_{xx} & 0 \\ 0 & G_{zz} \end{bmatrix}, \tag{7}$$

where $G_{xx} = G_x^T G_x$ and $G_{zz} = G_z^T G_z$, and

$$\begin{aligned} G_x &= D_x COS(\theta) - D_z SIN(\theta), \\ G_z &= D_z SIN(\theta) + D_z COS(\theta), \end{aligned} \tag{8}$$

where ϵ and δ are the Thomsen parameters, D_x and D_z represent the first order derivative operators with respect x and z respectively, G_x and G_z constitute the first order derivatives after axis rotation by the dip angle field θ, and $SIN(\theta)$ and $COS(\theta)$ are the corresponding trigonometric functions with their elements arranged as diagonal matrices. The corresponding transpose derivatives are given by

$$\begin{aligned} G_x^T &= COS(\theta)^T D_x^T - SIN(\theta)^T D_z^T, \\ G_z^T &= SIN(\theta)^T D_z^T + COS(\theta)^T D_z^T. \end{aligned} \tag{9}$$

Using Equations 8 and 9 we ensure that Equation 5 remains stable by keeping matrix $\mathbf{G}_2$ symmetric and negative definite (Zhang et al., 2011). In the current implementation I employ central differences and staggered grids for the derivative operators D_x and D_z, followed by linear interpolation to the original grid. Logically, D_x^T and D_z^T are implemented by applying the adjoint of the interpolation operator, followed by the adjoint of the derivative operator.

The definition of the TTI EIC beginning from the energy of the TTI wave equation (Equation 5) is similar to the VTI case (Cabrales-Vargas, 2016). In such a case the energy of the VTI wave equation is given by

$$E(t) = \int_\Omega \left\{ \frac{1}{v^2}[\|\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{Y}\|^2 \dot{\mathbf{p}} + \|\mathbf{D}\mathbf{p}\|^2 \right\} d\mathbf{x} = \int_\Omega \|\Box\mathbf{p}\|^2 d\mathbf{x}, \tag{10}$$

where the columns of matrix $\mathbf{Y}$ are the eigenvectors of matrix $\mathbf{B}$, and the diagonal of matrix $\mathbf{\Lambda}$ contains the eigenvalues of matrix $\mathbf{B}$. Such matrices are obtained from the diagonalization process: $\mathbf{B} = \mathbf{Y}^T \mathbf{\Lambda} \mathbf{Y}$. Matrix $\mathbf{D}$ represents the first derivative operator, defined as

$$\mathbf{D} = \begin{bmatrix} \frac{\partial}{\partial x} & 0 \\ 0 & \frac{\partial}{\partial z} \end{bmatrix}. \tag{11}$$

With these definitions the space-time gradient for the VTI wave equation becomes a matrix-like operator defined by

$$\Box = \left(\frac{1}{v} \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{Y} \frac{\partial}{\partial t}, \mathbf{D} \right) \tag{12}$$

and the modified version with cut-off angle

$$\widetilde{\Box} = \cos(2\gamma_c) \left(\frac{1}{v} \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{Y} \frac{\partial}{\partial t}, \mathbf{D} \right). \tag{13}$$

Therefore, the VTI EIC can be written as

$$\begin{aligned} I(\mathbf{x}; \gamma_c) = & \sum_t \Box \mathbf{p}_S \cdot \widetilde{\Box} \mathbf{p}_R = \qquad (14) \\ & \sum_t \left\{ \cos(2\gamma_c) \frac{1}{v^2} [\mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{Y} \dot{\mathbf{p}}_S] \cdot [\mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{Y} \dot{\mathbf{p}}_R] + [\mathbf{D} \mathbf{p}_S] \cdot [\mathbf{D} \mathbf{p}_R] \right\}. \end{aligned}$$

For TTI we substitute $\mathbf{G}$ for $\mathbf{D}$,

$$E(t) = \int_\Omega \left\{ \frac{1}{v^2} [\| \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{Y} \|^2 \dot{\mathbf{p}} + \| \mathbf{G} \mathbf{p} \|^2 \right\} d\mathbf{x} = \int_\Omega \| \Box \mathbf{p} \|^2 d\mathbf{x}, \tag{15}$$

where $\mathbf{G}$ is given as

$$\mathbf{G} = \begin{bmatrix} G_x & 0 \\ 0 & G_z \end{bmatrix} \tag{16}$$

The corresponding TTI EIC thus becomes

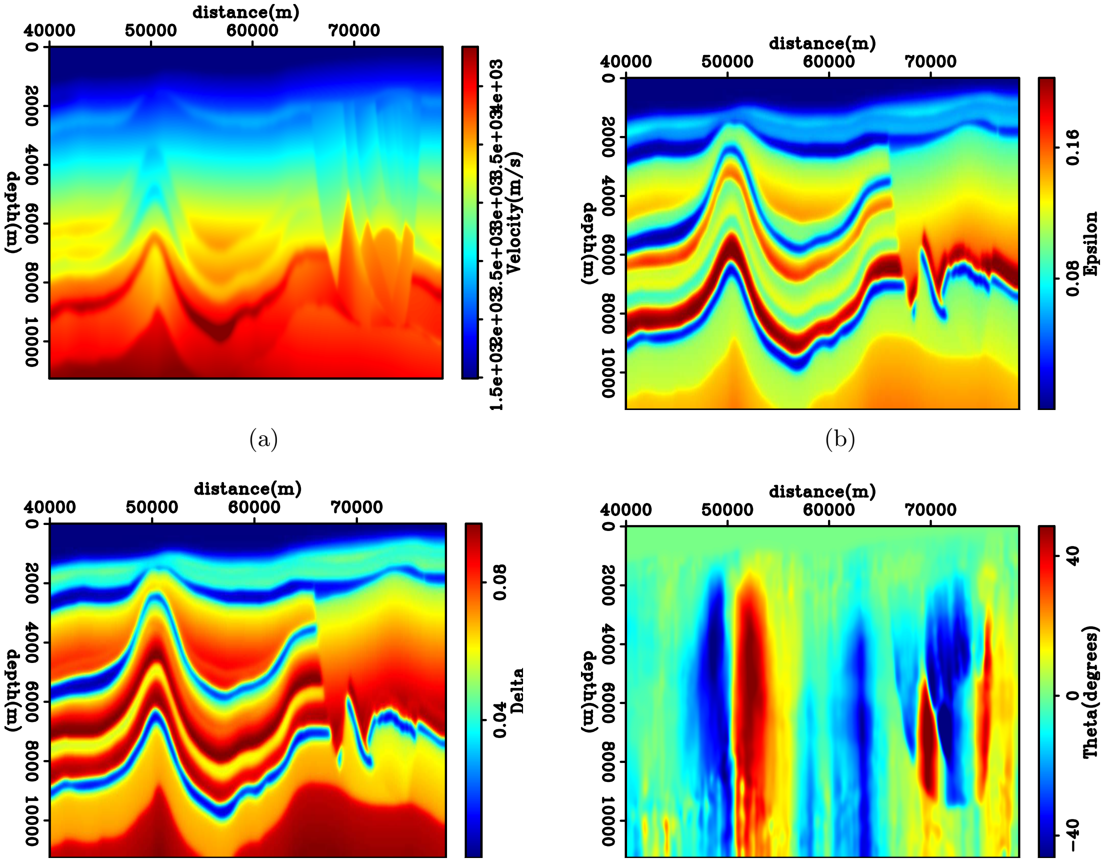
$$\begin{aligned} I(\mathbf{x}; \gamma_c) = & \sum_t \Box \mathbf{p}_S \cdot \widetilde{\Box} \mathbf{p}_R = \qquad (17) \\ & \sum_t \left\{ \cos(2\gamma_c) \frac{1}{v^2} [\mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{Y} \dot{\mathbf{p}}_S] \cdot [\mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{Y} \dot{\mathbf{p}}_R] + [\mathbf{G} \mathbf{p}_S] \cdot [\mathbf{G} \mathbf{p}_R] \right\}. \end{aligned}$$

The image normalization employed in the VTI case (Cabrales-Vargas, 2016) is applicable for TTI EIC. I use such normalization in the numerical examples presented in the next section.

SYNTHETIC EXAMPLES

I compare the result of the EIC with the crosscorrelation imaging condition (CIC) by applying anisotropic RTM to one half of the BP TTI synthetic dataset (Figure 1(a)).

Figure 2 shows the TTI-RTM result using the CIC. The seismic events are obscured by the tomographic component of the RTM image. Applying the EIC with $\gamma_c = 0°$ (Figure 3) we can preserve such tomographic component. Figures 2 and 3 are very similar, but closer examination reveals that in fact, using CIC the reflections are simply *obscured*, while using EIC for $\gamma_c = 0°$ the reflections become *attenuated*. Figures 4 and 5 show the corresponding



(a) (b)

(c) (d)

Figure 1: BP TTI synthetic model, after smoothing: (a) Velocity; (b) Epsilon; (c) Delta; (d) Theta (dip) [**ER**]. alejandro2/. vmodel,emodel,dmodel,tmodel

sections after the application of a Laplacian filter, confirming the last argument. The oceanic floor reflector persists because it encompases a wide reflection angle range, whereas the cut-off angle has the rejection band confined to a small threshold, in this case around 0°. Rocha et al. (2016a) discuss the potential benefit of attenuating the reflectivity while preserving the tomographic component in the computation of the full-waveform inversion gradient, thus reducing the susceptibility to cycle skipping.

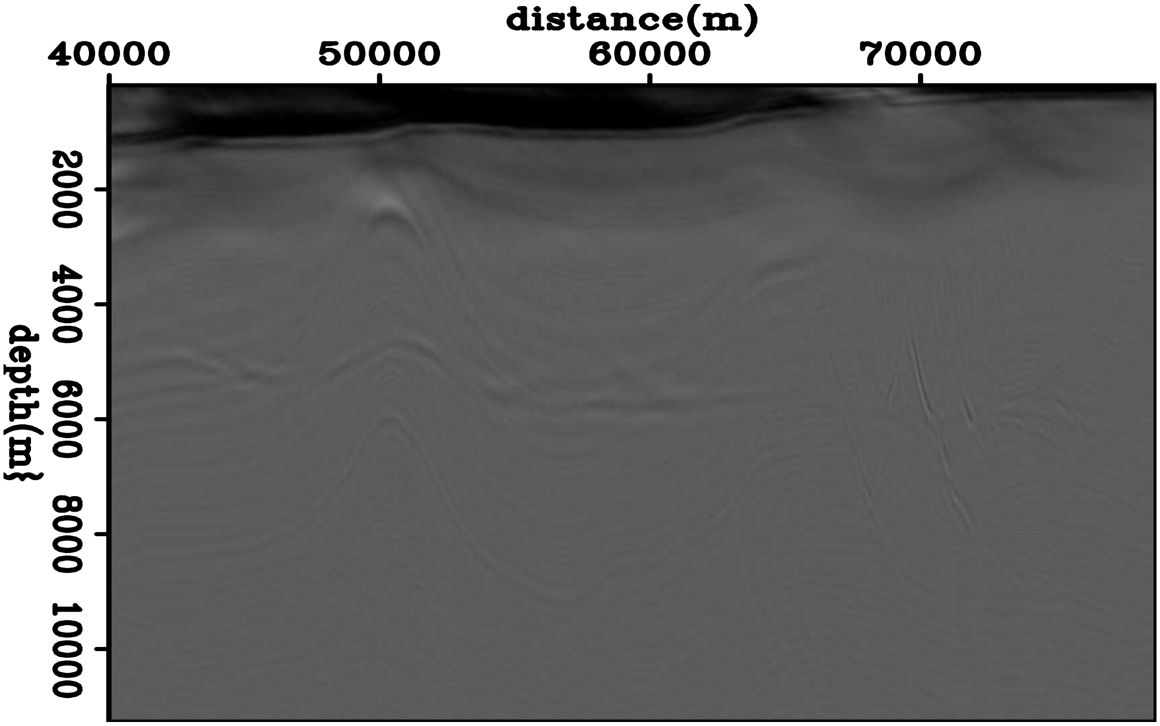


Figure 2: TTI reverse-time migration using CIC [**CR**]. alejandro2/. cic

Figure 6 shows the result of the EIC setting $\gamma_c = 90°$ to preserve reflectivity, to be compared with Figure 4. This EIC section still exhibits remaining artifacts in the shallow part, which are absent in the filtered version of CIC. On the other hand, the Laplacian filter slightly affects the continuity of some shallow events near the ocean bottom. In contrast to the acoustic case (Rocha et al., 2015, 2016a), it is possible that in anisotropic propagation, γ_c values other than 90° might give better results in attenuating the tomographic component. In fact, the EIC offers the flexibility to choose among different rejection angles, but additional research is needed to determine the optimum value of γ_c without relying on trial and error.

CONCLUSION

I derived a TTI energy imaging condition from the VTI case presented in the previous SEP report. The numerical results show satisfactory preservation of the tomographic component when setting $\gamma_c = 0°$. Preservation of the reflectivity component was less than optimum, possibly because the anisotropic case might require to set γ_c to angle values other than 90°. Therefore, additional research is needed to determine the optimum cut-off angle.

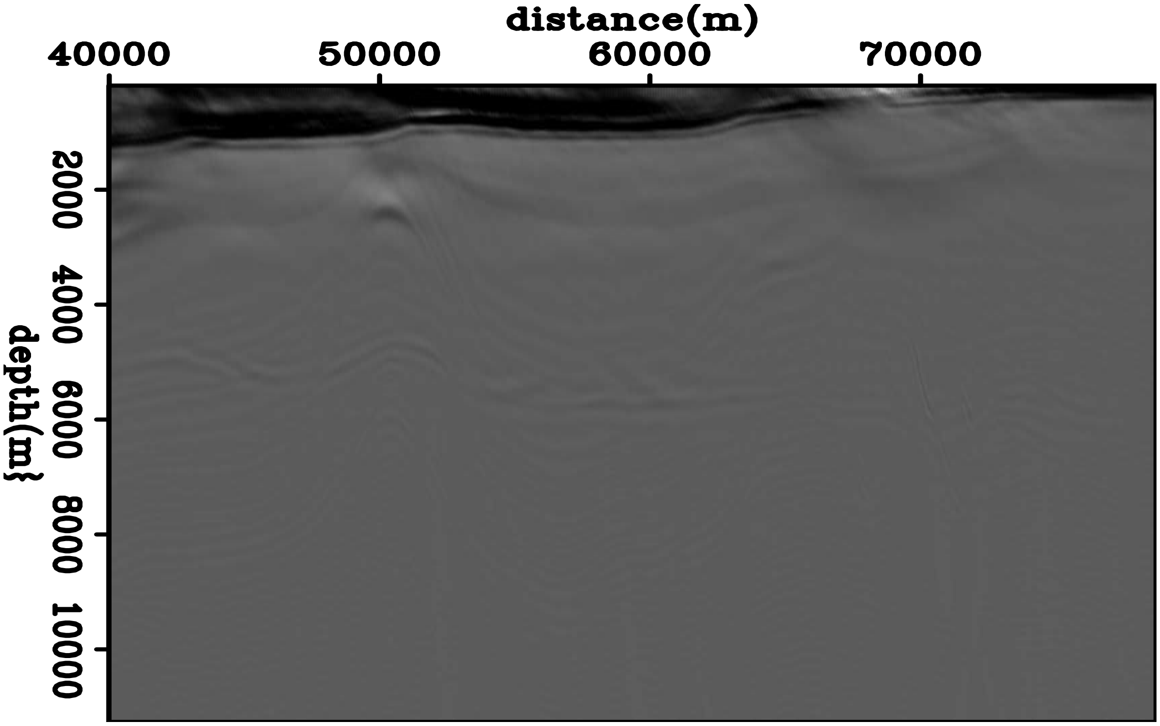


Figure 3: TTI reverse-time migration usinf EIC with $\gamma_c = 0°$ [**CR**]. alejandro2/. eic1

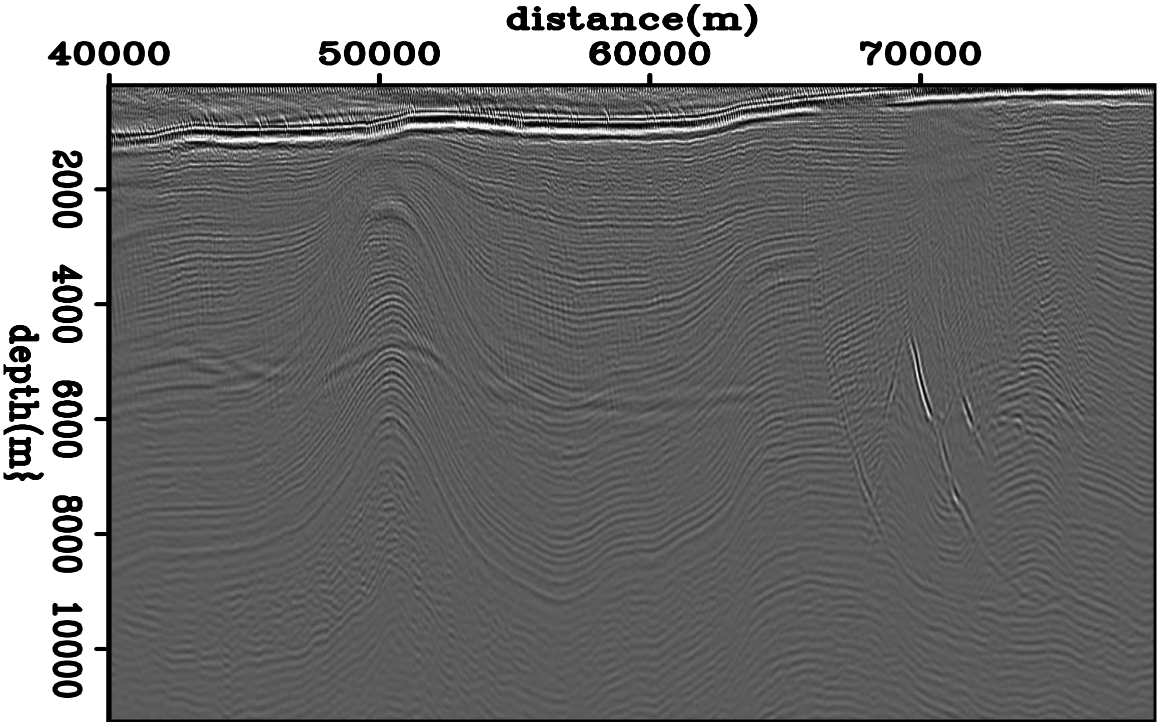


Figure 4: TTI reverse-time migration using CIC, after the application of a Laplacian filter [**CR**]. alejandro2/. cic-lc

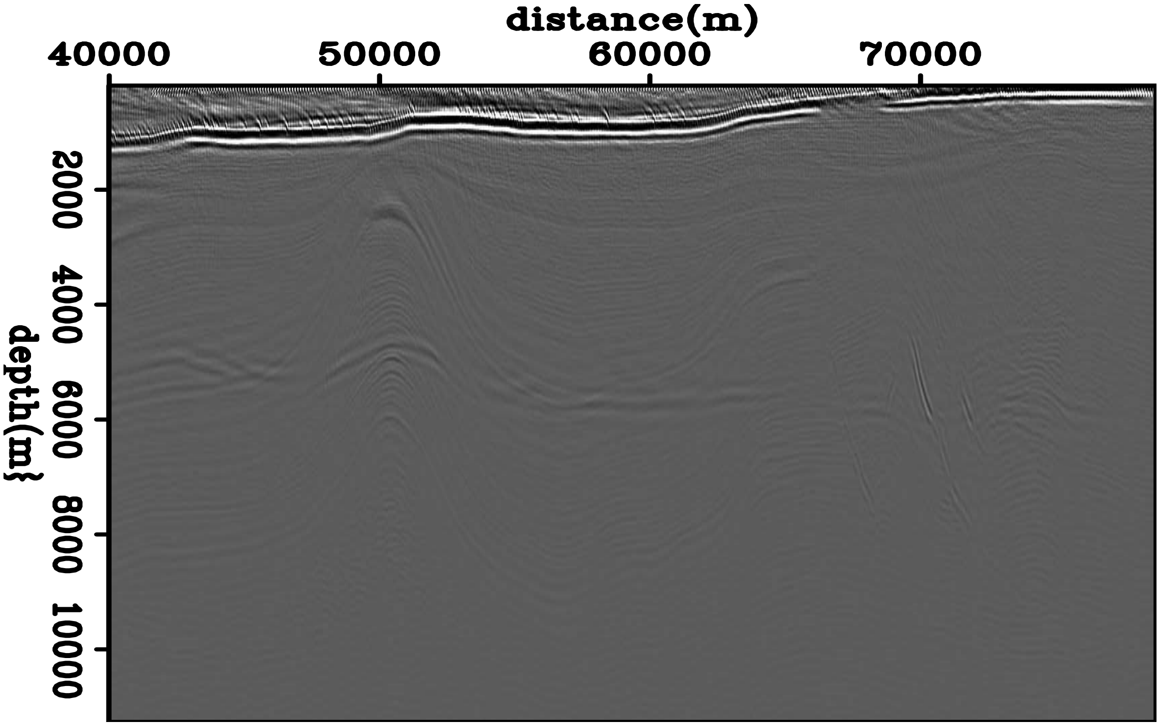


Figure 5: TTI reverse-time migration usinf EIC with $\gamma_c = 0°$, after the application of a Laplacian filter [**CR**]. alejandro2/. eic1-lc

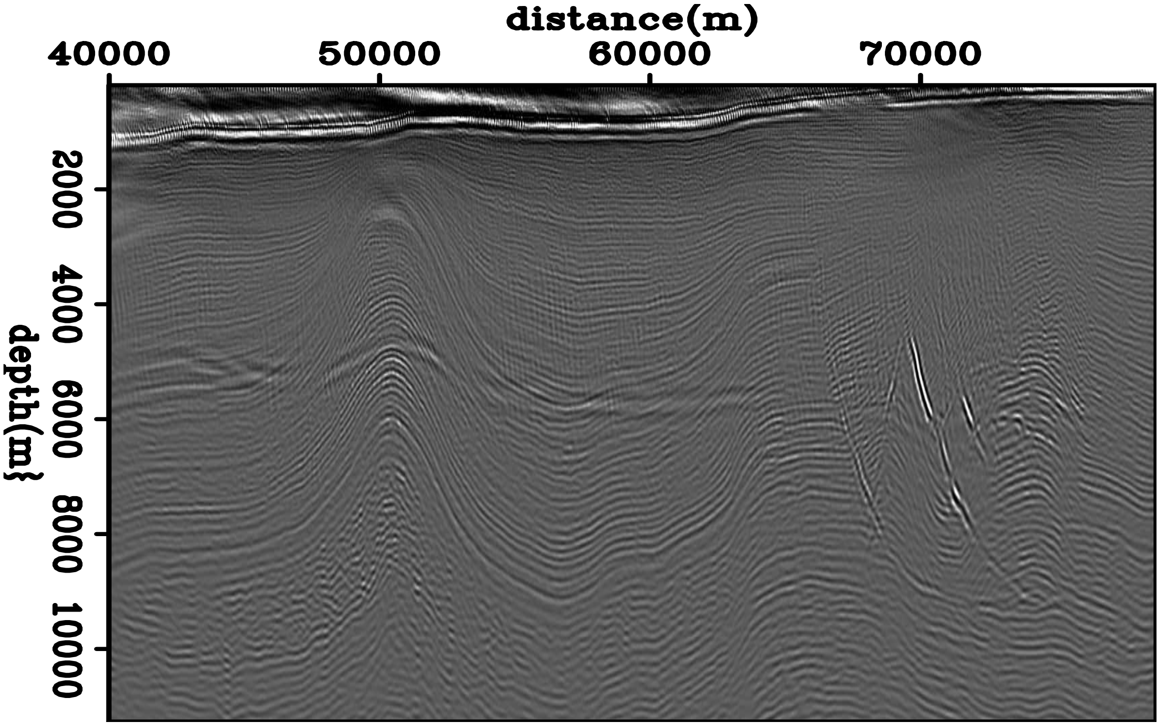


Figure 6: TTI reverse-time migration usinf EIC with $\gamma_c = 90°$ [**CR**]. alejandro2/. eic2

ACKNOWLEDGEMENT

I would like to thank the SEP sponsors for their support, and to British Petroleum for the synthetic BP TTI model. I also extend special thanks to Petróleos Mexicanos for the financial support.

REFERENCES

Cabrales-Vargas, A., 2016, Reverse-time migration using the energy imaging condition in isotropic and vti media: SEP-Report, **163**, 77–90.

Rocha, D., N. Tanushev, and P. Sava, 2015, Acoustic wavefield imaging using the energy norm: SEG Technical Program Expanded Abstracts 2015, 4085–4090.

———, 2016a, Acoustic wavefield imaging using the energy norm: Geophysics, **81**, no. 4, S151–S163.

———, 2016b, Isotropic elastic wavefield imaging using the energy norm: Geophysics, **81**, no. 4, S207–S219.

Whitmore, N. and S. Crawley, 2012, Applications of RTM inverse scattering imaging conditions: SEG Technical Program Expanded Abstracts 2012, 1–6.

Zhang, Y., H. Zhang, and G. Zhang, 2011, A stable TTI reverse time migration and its implementation: Geophysics, **76**, no. 3, WA3–WA11.

Stanford Exploration Project, SEP165, October 8, 2016

Application of wave-equation migration velocity and Q analysis to the field data from the North Sea

Yi Shen

ABSTRACT

The Dolphin Geophysicals (Dolphin) multi-client field data acquired in the North Sea used in this study has attenuation problems. The area was under the influence of salt tectonics, producing two diapirs. Dolphin interpreted a gas chimney above one diapir, and a channel above the other. The gas chimney forms a migration pathway for the gas to leak and then to accumulate at the shallow position. The shallow gas gives rise to strong attenuation and low interval velocities in the gas area. The channel also has low velocities, and strong attenuation is associated with it. The objective of my study is to update the provided velocity model, especially at the gas and channel area, and to invert for the Q models to recover these two anomalies. Angle domain common image gathers after migration with the current interval velocity show that most of the events are curved down, indicating the current velocity is too high. Furthermore, both the migrated image and the angle gathers show that the events between 26,000 meters (m) to 28,000 m and 38,000 m to 42,000 m are wiped out below the two salt bodies. The attenuation anomalies above two salt bodies are the main reasons for the wiped-out image below. I first applied wave-equation migration velocity analysis to update the current velocity model. As a result, the velocity decreases in the gas and channel area. The angle gathers migrated using the updated velocity model are much more flattened, and the events above the top of salt in the migrated images are more coherent. Then, I applied wave-equation migration Q analysis to invert for the Q models. The estimated Q model shows that the two Q anomalies are recovered and match the interpretation. By using this Q model in seismic migration, I made the seismic events below the anomalies clearly visible, with improved frequency content and coherency of the events.

INTRODUCTION

Seismic attenuation, typically quantified by a parameter Q, is a notoriously challenging problem for reservoir identification and interpretation in the North Sea, where strong attenuation anomalies are present. Attenuation degrades the seismic image quality by decaying the image amplitude, lowering the image resolution, distorting the phase of events, and dispersing the velocity. These problems impede accurate image interpretation for hydrocarbon production and well positioning. The Dolphin's multi-client field data acquired in the North Sea (CNS data) used in this study has such an attenuation problem. Gas chimneys and channels exist in the subsurface with strong attenuation and low interval velocity. These complexities reduce the amplitude and phase of deeper events, and essentially create a shadow zone over the salt body below and at the potential reservoirs, thus hampering accurate reservoir interpretation. Therefore, it is important to build a velocity model as accurate as possible. It is also valuable to understand and quantify the effects of these attenuation anomalies to create an accurate laterally- and vertically- varying attenuation model. The improvements in image quality using the derived model provide greater confidence for hydrocarbon exploration.

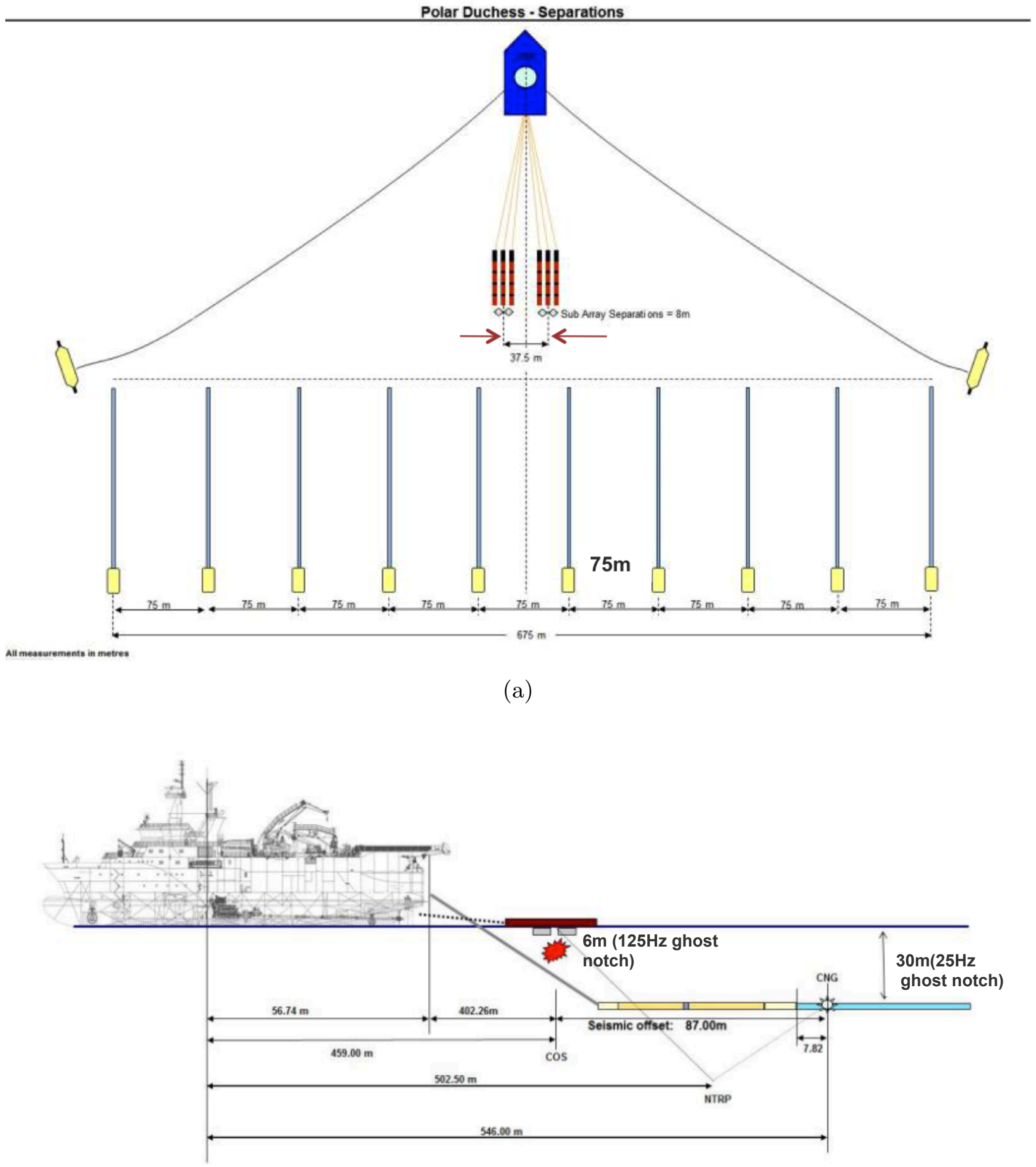
In my previous studies, I have shown the methodology and workflow to update the Q model using wave-equation migration Q analysis (WEMQA) (Shen et al., 2013, 2014, 2015; Shen* et al., 2015). I have also included the velocity inversion using wave-equation migration velocity analysis (WEMVA) (Sava and Biondi, 2004; Zhang and Biondi, 2013) into the Q inversion workflow to update both the velocity and Q models (Shen, 2015). I have demonstrated the successful application of WEMQA combined with WEMVA on a modified synthetic SEAM model (Shen, 2015). In this study, I continuously use this workflow (Shen, 2015) to build velocity and Q models for this field data to recover the gas and channel; and therefore to produce an improved seismic image.

In this paper, I first give an overview of these field data. Second, I present my preprocessing workflow to prepare the data for the later analysis. Third, I apply WEMQA combined with WEMVA to update the current velocity model and to invert for the Q models to recover the Q anomalies.

DATA DESCRIPTION

The CNS data is acquired using a marine-towed streamer seismic survey in the North Sea. The data provided by Dolphin for this research contain 28 sail lines, which were acquired by three vessels. The streamer configurations for the survey of three vessels are similar. Figure 1(a) shows the streamer configuration for the survey of one of the vessels. This vessel has 10 streamers. Each streamer is 6,000 m long, and the separation between the neighboring streamers is 75 m. The source was configured with two shots, in a flip-flop mode with a shot interval of 25 m. The source separation was 37.5 m. The streamer depth for all the survey was 30 m, and the source depth was 6 m, as shown in Figure 1(b). As a result, both the receiver-side ghost notch and source-side ghost notch are able to be calculated using equation $f = c/2d$, where f is the frequency of the first ghost notch, c is the water velocity of $1500m/s$ and d is the streamer/source depth. The results show that the receiver-side ghost has its first notch at 25 Hz, and the source-side ghost has its first ghost notch at 125 Hz.

The CNS data have strong attenuation anomalies. The depth slice provided by Dolphin in Figure 2 highlights the areas with the strongest gas chimneys and strongest channel effect. The depth slice in Figure 2 does not cover the entire survey, but it covers the most complex area for the study of these anomalies. According to the interpretation by Dolphin, attenuation appears stronger from gas chimneys than from the channels. Attenuation from the gas on the right dome of the slice is stronger than the one on the bottom left dome of the slice. The channel at the top left of the slice shows low velocities, and there might be Q anomalies with it, although exhibiting less attenuation. The part of the survey provided by Dolphin for this study covers the left gas chimney and channel but does not include the anomaly on the right of the slice.



(a)

(b)

Figure 1: (a) The streamer configuration for the survey and (b) the boat configuration of one of the three vessels that are involved in these field data acquisition. [**NR**] yishen1/. dlp-boat3,dlp-boat2

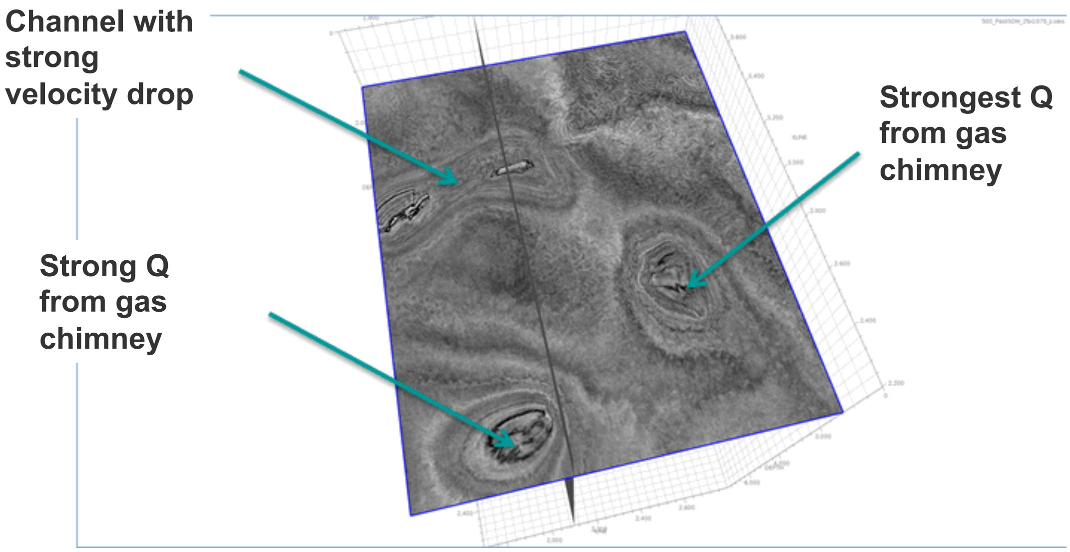


Figure 2: A depth slice provided by Dolphin that highlights the areas with the strongest anomalies associated to gas chimneys and channel. North is upward. [**NR**] yishen1/. dlp-depth

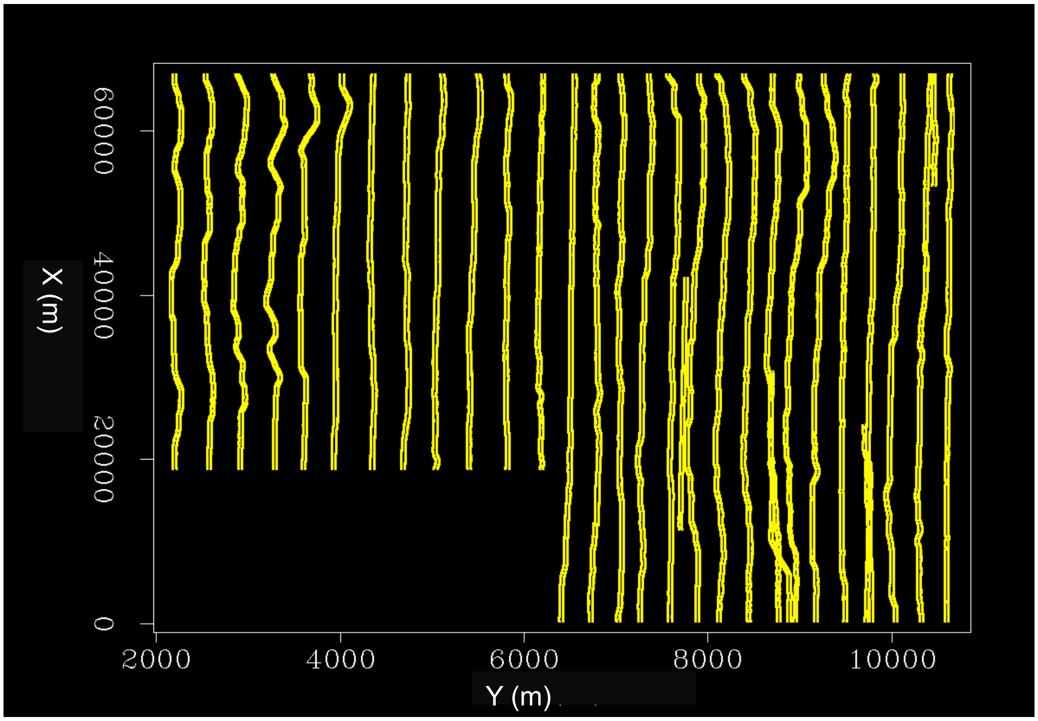
PREPROCESSING

Coordinates manipulation

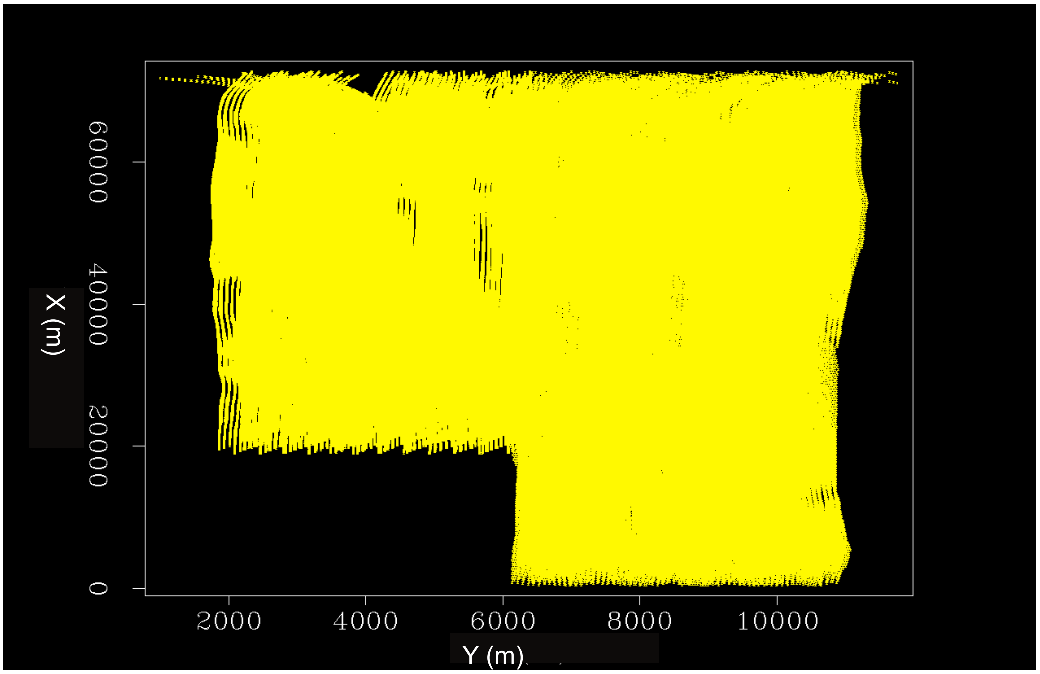
The original coordinates of source and receiver of CNS data provided by Dolphin are not regularly sampled along neither inline nor crossline direction, and they are in a rotated Cartesian coordinate system. First, I shifted the origins of the coordinates and rotated the tilted coordinates to align the X–axis with the inline direction and the Y–axis with the crossline direction. The rotated and shifted coordinates of source and receivers are shown in Figure 3. Aligning the processing grid axes with the acquisition inline and cross-line directions makes it easy to bin the seismic shot gathers to the regularly sampled data grid that I create. Then, I regularized the grids of the source and receiver coordinates using the parameters shown in Table 1, facilitating subsequent data processing.

Table 1: The parameters used for regularizing the grids

Keys	number	origin [m]	spacing [m]
Receiver at crossline (Y)	10	-337.5	75
Receiver at inline (X)	240	50	25
Source at crossline (Y)	28	2100	300
Source at inline (X)	1239	230	50



(a)



(b)

Figure 3: Rotated and shifted coordinates of (a) shots and (b) receivers. [**NR**] yishen1/. dlp-src-rot2,dlp-rec-rot2

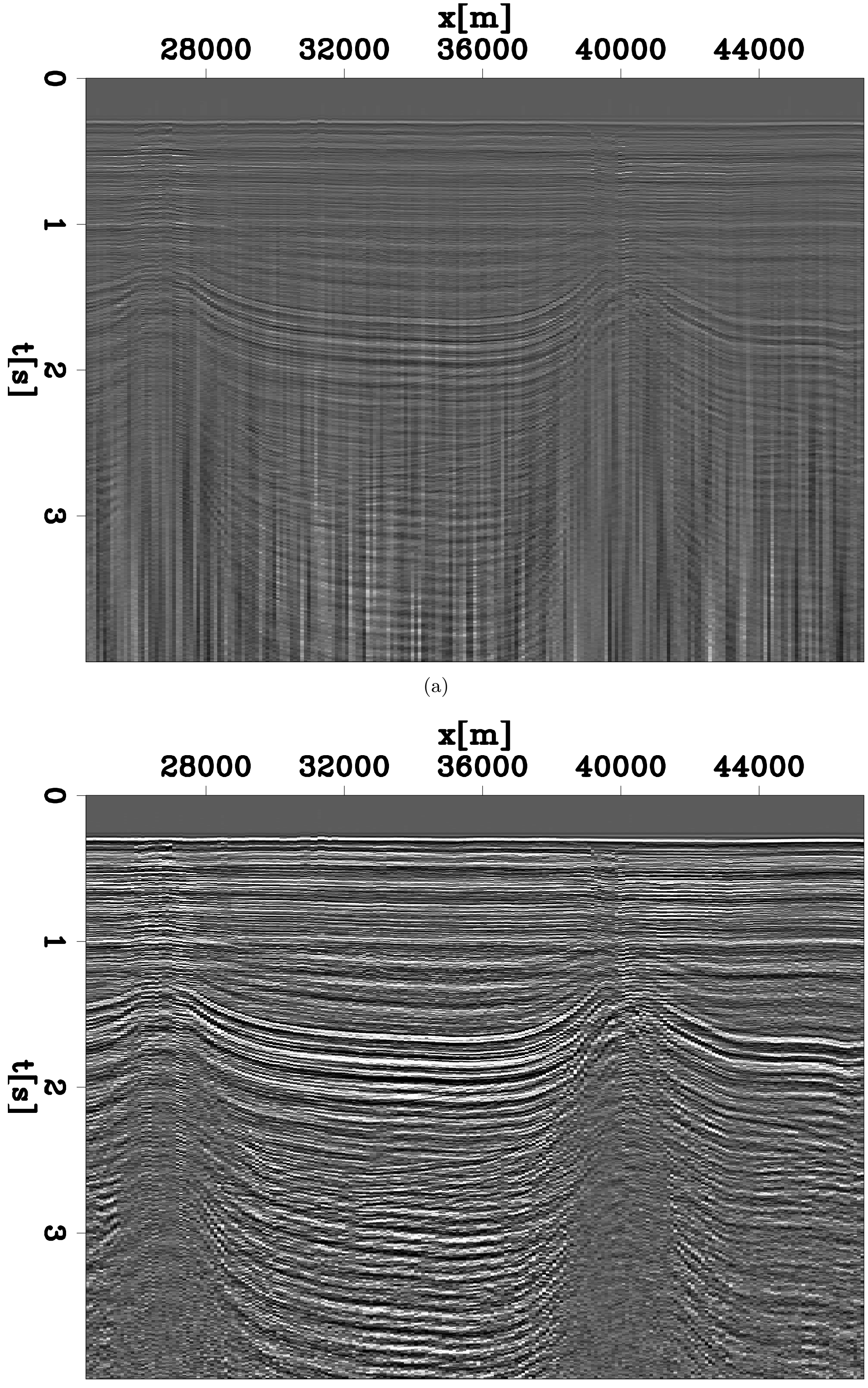
Denoise, Demultiple, Debubble and Deghost

This data is strongly influenced by swell waves that propagate along the interface between water and air. Such noise can also be identified in the FK domain of the data. Figure 4(a) shows a common offset gather at offset=300 m. The vertical strips in the common offset gather are the low-frequency noise that is generated by the swell noise. Figure 5(a) shows the same common offset gather of Figure 4(a) in the FK domain. The horizontal strip around 0 Hz frequency corresponds to the low-frequency noise. Therefore, I use a low-cut filter at 2 Hz to remove such noise. Figure 5(b) shows the common offset gather in the FK doamin after filtering. The result in the physical domain is shown in Figure 4(b). Notice the virtual absence of vertical strips, which means the swell noise is removed.

Beside the swell noise, these data are also contaminated by water-bottom multiples, salt-interval multiples, bubbles from the airguns, and the ghost described in the previous section. These noises copy events, e.g. water bottom reflection, at a periodic spatial interval. I used a gapped Preconditioning Error Filter (PEF) to remove such repetitive patterns (Claerbout and Fomel, 2014). Figure 6(a) and 6(b) are the windowed common offset gather at offset=300 m before and after a gapped PEF being applied, respectively. The yellow arrows show the bubbles, ghosts and multiples in Figure 6(a). Such events, of which the locations are also indicated by yellow arrows in Figure 6(a), are partially removed. Figure 7(a) and Figure 7(b) show a representative shot gather before and after a gapped PEF being applied, respectively. Figure 7(a) shows the bubbles, ghosts and multiples that are pointed by the yellow arrows. The event marked as repetitive event is still unknown to me, which may be a result of a wave bouncing from the back of the boat. The locations of these marked events are removed from Figure 7(b) after a gapped PEF being applied. The events The spectra of Figure 7(a) and Figure 7(b) are compared in Figure 8. The first receiver-side ghost notch is flattened by the preprocessing with a gapped PEF. The noisy wiggles can be observed at the high frequencies in Figure 8, because PEF tries to flatten the spectra and therefore boosts the high frequency noise.

TWO-DIMENSIONAL VELOCITY AND Q ESTIMATION

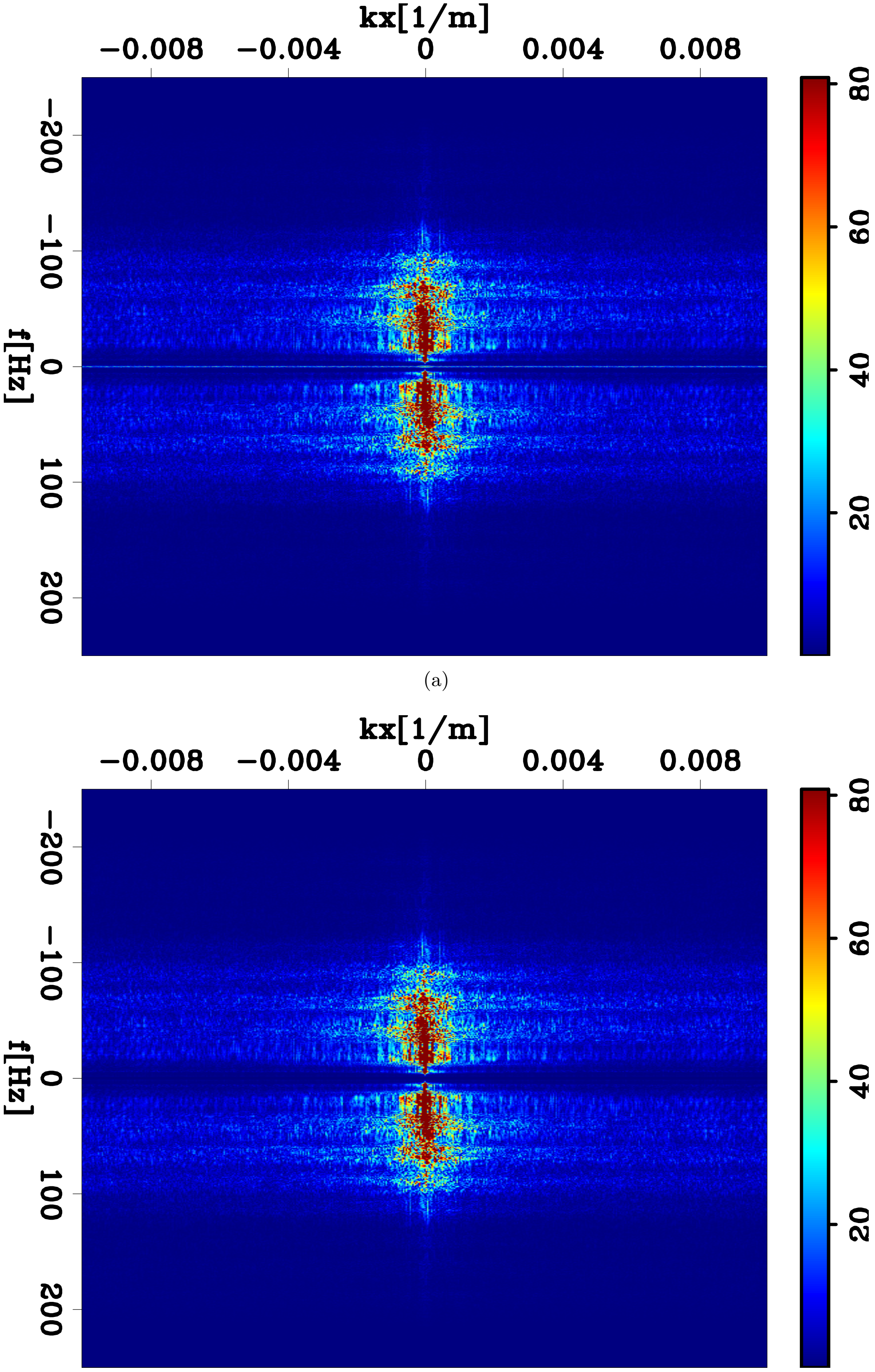
For the model building for this field study, I first focused the analysis on one representative two-dimensional (2D) section in order to set a proper inversion workflow and the corresponding parameter set for this particular dataset, and thus better prepare the road to 3D field inversion. In this report I show the results for the 2D inversion only. This 2D section is at the crossline 7,500 m, which passes through the left-side gas chimney and the left channel as shown in Figure 2. Figure 9 is the 2D slice of the 3D depth NMO interval velocity model provided by Dolphin at the crossline of 7,500 m, as an initial velocity model for the inversion. The velocity model shows a salt body with two diapirs, and a high-velocity layer above the salt body. Such layer acts as a potential cap layer. A gas chimney constitutes a migration pathway for the trapped gas that leaks and accumulates at the shallow position above the left diapir. The shallow gas presents low velocity and strong attenuation. Therefore, the velocities shown in Figure 9 at the shallow position above the left salt domes are slower than the surrounding areas. The area above the right salt domes also has slow velocities as shown in Figure 9, because of the presence of a low-velocity channel that as interpreted by Dolphin. Because Dolphin has not provided the Q model, the initial Q model for the inversion is set to be homogeneous with a value of 500.



(a)

(b)

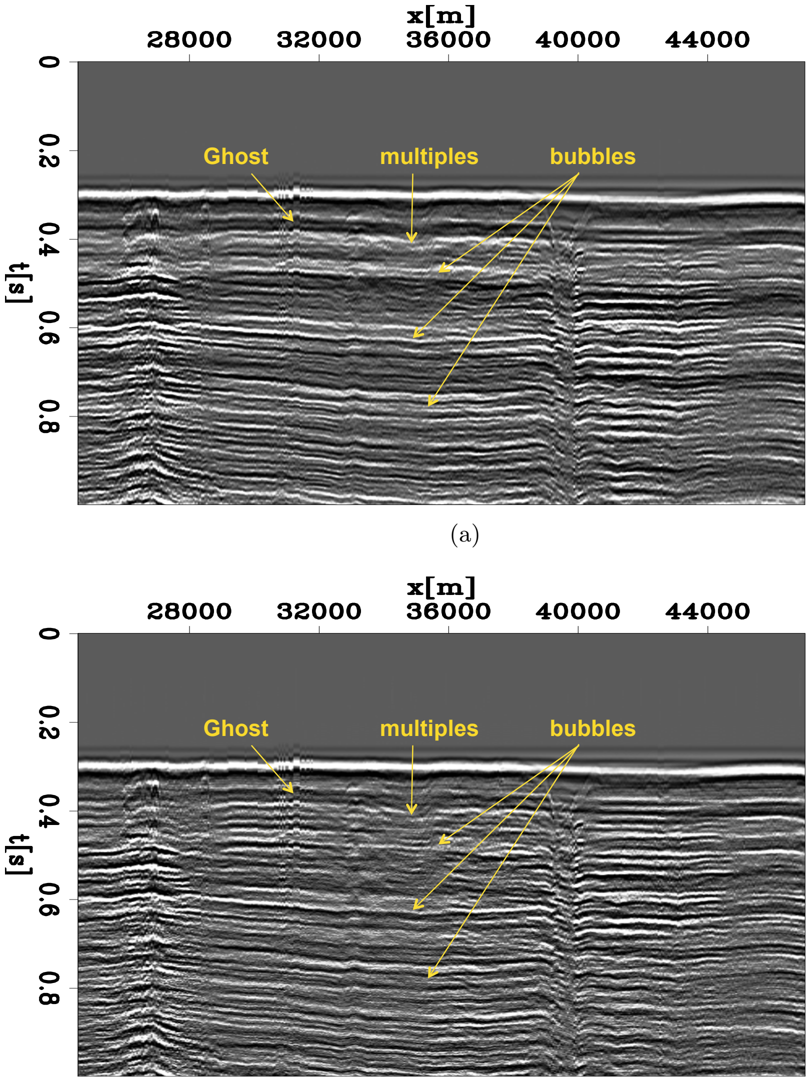
Figure 4: Common offset gather at offset=300 m: (a) before denoise; (b) after denoise. [**CR**] yishen1/. dlp-dprc-offh,dlp-mprc-offh-fkd



(a)

(b)

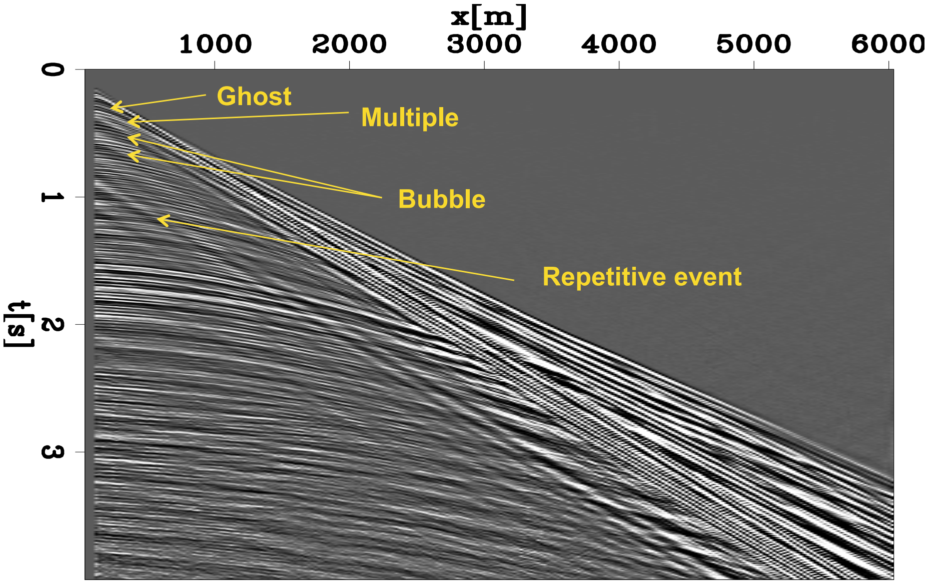
Figure 5: FK domain transformed from the common offset gather at offset=300: (a) before low-cut filter being applied; and (b) after low-cut filter being applied. [**CR**] yishen1/. dlp-mprc-offh-fkb,dlp-mprc-offh-fka



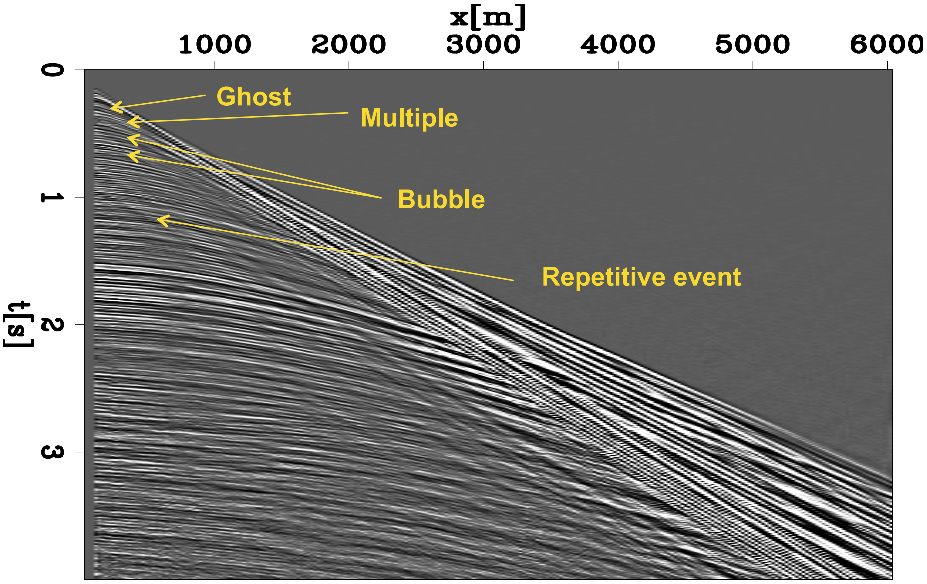
(a)

(b)

Figure 6: Windowed common offset gather at offset=300 m: (a) before processed with a gapped PEF; (b) after processed with a gapped PEF. The marked events are removed. [**CR**] yishen1/. dlp-mprc-offh-fkd-win2,dlp-mprc-offh-pefm3-win2

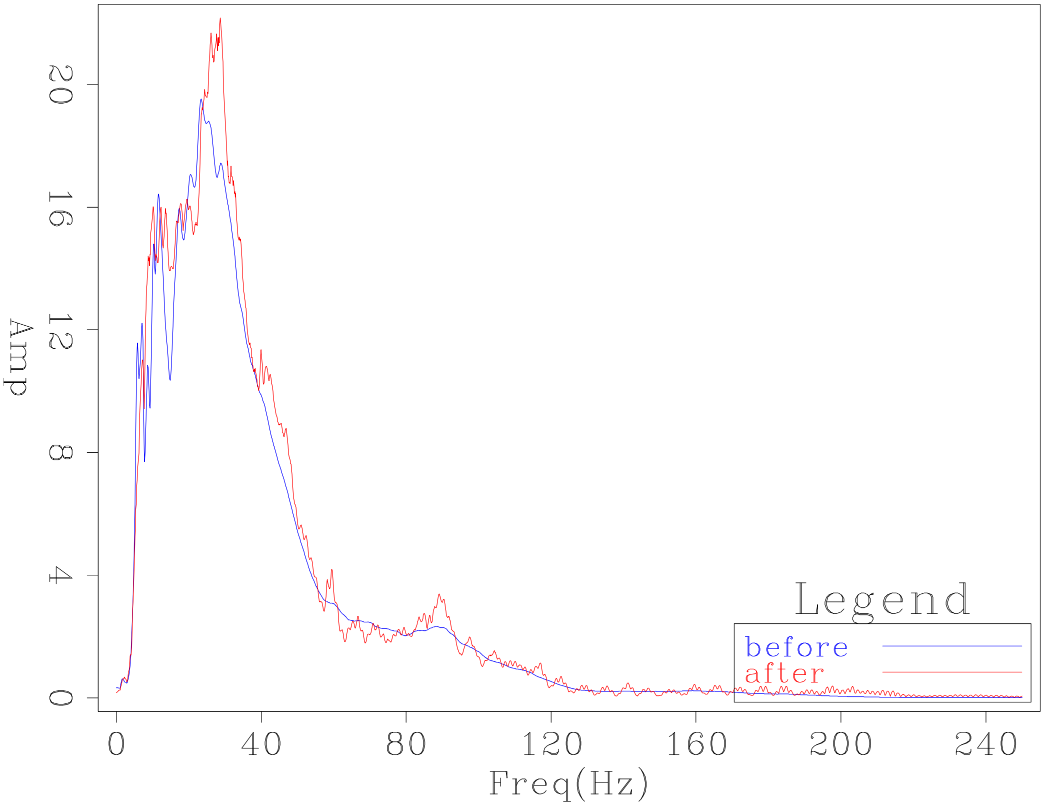


(a)



(b)

Figure 7: Shot gather at inline=24,580 m, crossline =7,500 m: (a) before processed with a gapped PEF; (b) after processed with a gapped PEF. The marked events are removed. [**CR**] yishen1/. dlp-mprc-shte-fkd2,dlp-mprc-shte-pefm3



(a)

Figure 8: The spectra of Figure 7(a) and Figure 7(b). [**CR**] yishen1/. dlp-mprc-shte-spex

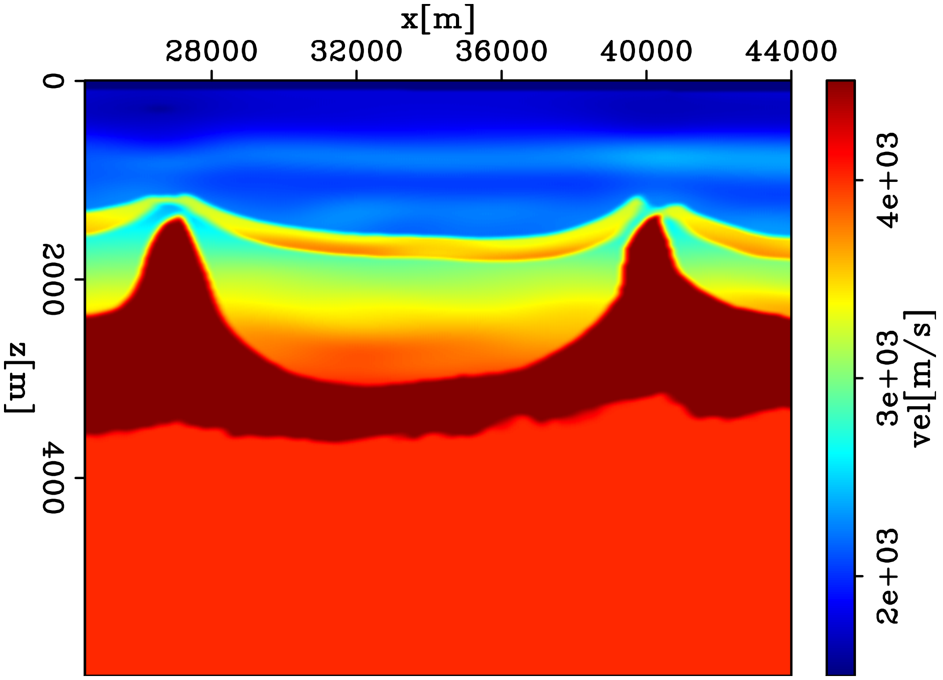


Figure 9: A 2D slice of the 3D depth NMO interval velocity model provided by Dolphin at the crossline of 7500 m, which passes through the left-side gas chimney and the left channel as shown in Figure 2. [**CR**] yishen1/. dlp2D-bqiv-post-basc-bvel-iter0

Figure 10 is the migrated image using one-way wave-equation migration (Shen et al., 2013, 2014) at zero subsurface offset using the initial velocity model and the initial Q model. The migration frequency range is from 0.8 Hz to 50 Hz. The spacing of the imaging grid used is 25 m in X by 25 m in Y by 10 m in Z. The salt body with two domes ($x = 26,000$ m to $28,000$ m and $x = 38,000$ m to $42,000$ m) is well imaged. The bright spots above the left and right salt peaks are the shallow gas and channel, respectively. These two regions have strong sand-shale impedance contrasts and thus their seismic reflection amplitudes are strong, which results in bright spots. The events below these two spots are wiped out. The attenuation caused by the shallow gas and channel is the main reason for the dimming region below. The dimming image at the boundary ($x < 25,000$ m and $x > 43000$ m) are tapered by the boundary condition in the migration. Figure 12 shows 10 representative Angle Domain Common Image Gathers (ADCIGs) of this migrated image using the initial models. The events in the ADCIGs are not flat but curved down, thereby indicating the velocity is too high. Angle gathers with midpoints close to the attenuation anomalies, e.g., ADCIGs at $x = 27$ km and $x = 41$, have lower amplitudes.

I ran the velocity and Q inversion using the workflow described in my previous work (Shen, 2015) for 40 iterations. The optimization scheme described by Shen (2015)contains two objective functions: (1) one mainly for velocity updates, and (2) the other mainly for Q updates. Such scheme employs a weighting parameter β to balance these two functions. The previous study (Shen, 2015) shows that an error in velocity can cause significant error in the Q inversion results. The conspicuously curved-down events in the ADCIGs (Figure 12) indicate a significantly velocity error, therefore I focus on the velocity updating at the

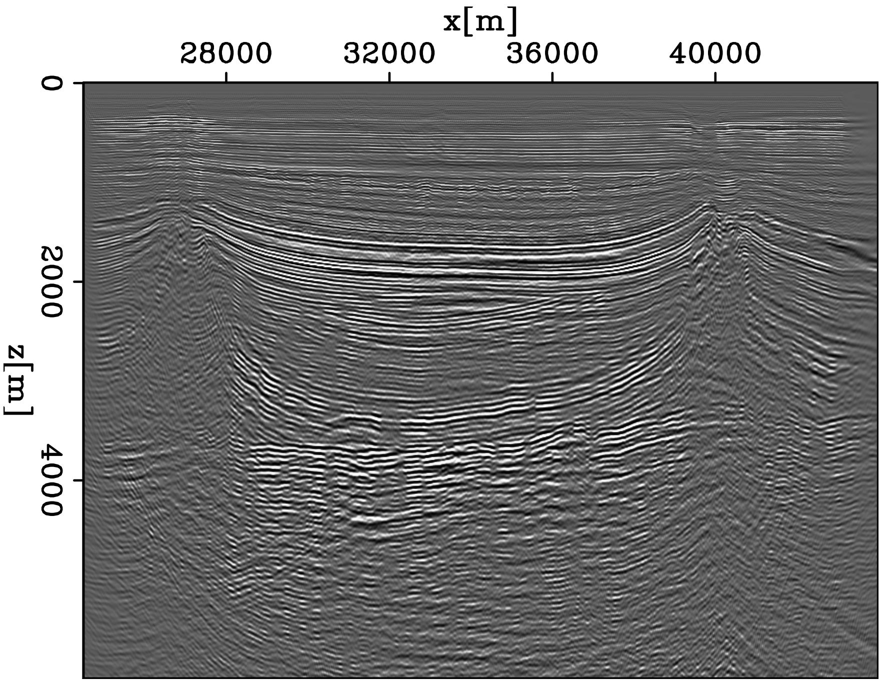


Figure 10: The migrated image at zero subsurface offset using the initial velocity model and the initial Q model. [**CR**] yishen1/. dlp2D-bqiv-post-basc-bimg-before

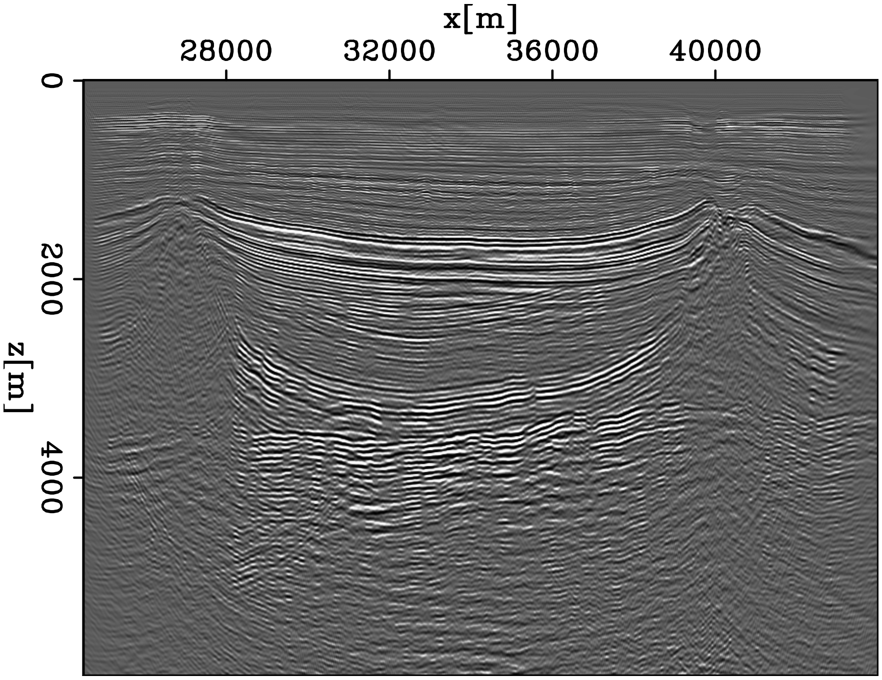


Figure 11: The migrated image at zero subsurface offset using the updated velocity model and the initial Q model. [**CR**] yishen1/. dlp2D-bqiv-post-basc-bimg-after

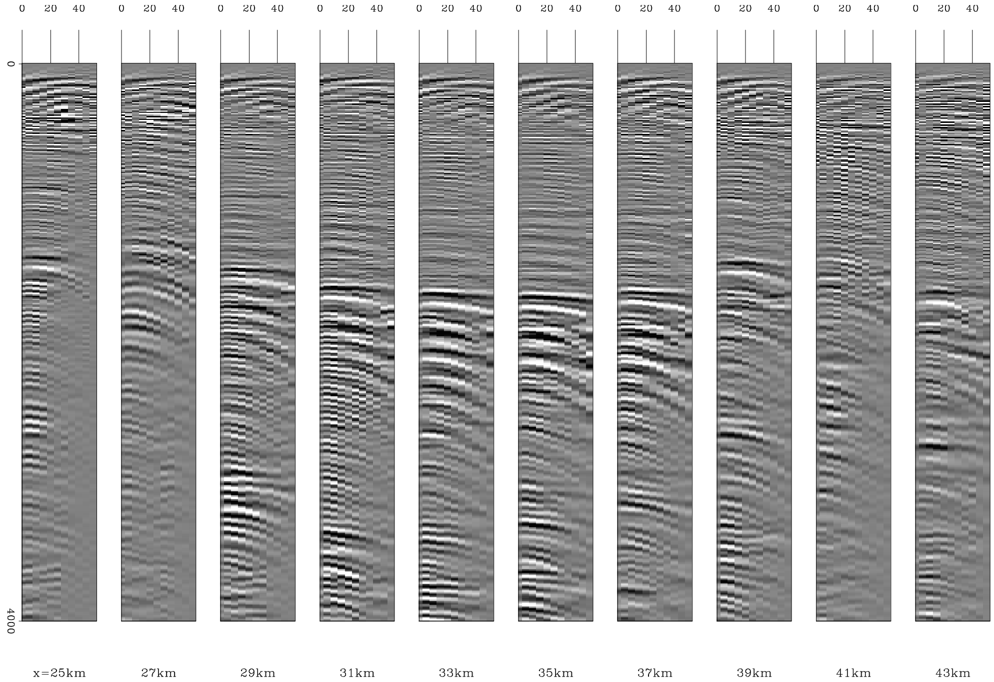


Figure 12: The ADCIGs using the initial velocity model and the initial Q model. [**CR**] yishen1/. dlp2D-bqiv-post-basc-bang-before

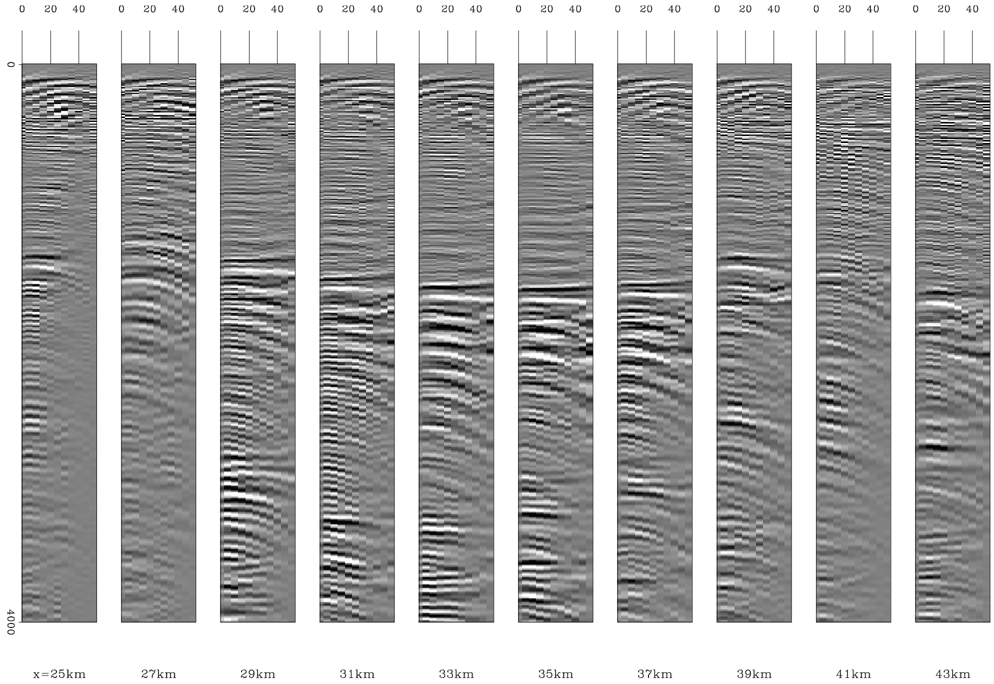


Figure 13: The ADCIGs using the updated velocity model and the initial Q model. [**CR**] yishen1/. dlp2D-bqiv-post-basc-bang-after

early iterations until the velocity error is significantly reduced. I update the Q model at the later iterations. In this inversion, I set the weighting parameter β to be small (10^{-10}) to mainly estimate velocity during the first 20 iterations, and set β to be large (10^{10}) to mainly focus on Q estimation during the final 20 iterations.

I chose stack power of the migration image as the objective function for velocity updates. This method maximizes and therefore focuses the energy at the zero subsurface offset of the migrated image. Its low computational cost makes the application feasible for the later 3D inversion running in the computer resources of academic institutes. Figure 14 shows the velocity gradient of the first iteration of the inversion, in which the sign is opposite of the searching direction. The gradient shows the updates are strong around the shallow gas and the channel region and aims to decrease the velocity. Figure 15 is the updated velocity after 20 iterations. Compared to the initial velocity in Figure 9, the shallow velocity becomes slower, and the shallow low-velocity zone (above 500 m) becomes thicker. The decreases of the shallow velocity intend to push the events in ADCIGs upward and therefore flatten those events. The velocity drops more significantly above the salt domes, as shown in Figure 15, corresponding to the low-velocity anomalies of shallow gas and channel.

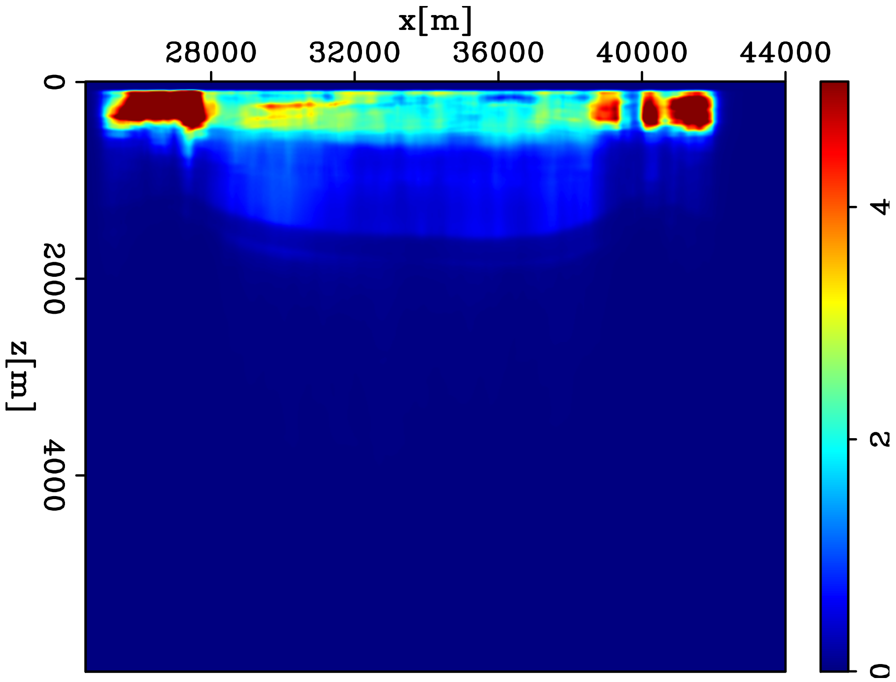


Figure 14: The velocity gradient of the first iteration of the inversion. [**CR**] yishen1/. dlp2D-bqiv-post-basc-grdp-iter0

Figure 11 is the seismic image at zero subsurface offset migrated using the updated velocity model and the initial Q model. The updates lower the velocity at the shallow part, and as a result, the events in Figure 11 move up. Figure 11 also shows how some events (e.g. the flank of the right salt dome) become more coherent in phase after velocity updating. Figure 13 shows 10 representative ADCIGs of this migrated image using the updated velocity model and the initial Q model. The decrease in the shallow velocity after inversion pushes the curves of the events in ADCIGs upward, and therefore flattens the gathers. Such flatness is significant above $z = 4,000$ m. More iterations or a different

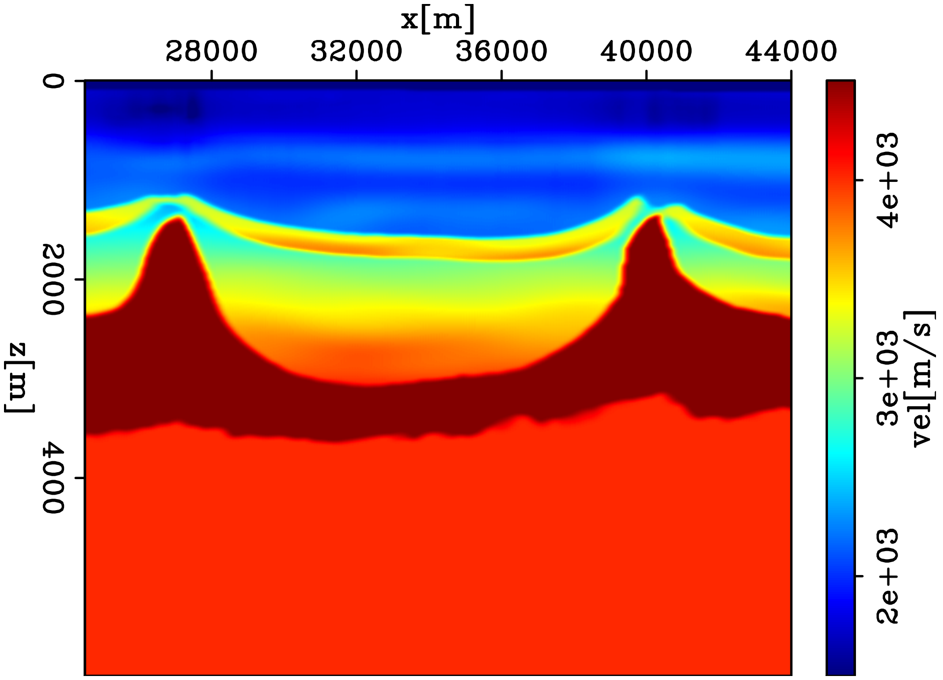


Figure 15: The updated velocity after 20 iterations. [**CR**]
yishen1/. dlp2D-bqiv-post-basc-bvel-iter20

objective function may improve the flatness in a deeper depth; however, events below the base of salt ($z > 4,000$ m) are not the target of my study, and have little (if any) impact on the estimation of the shallow gas and channel for both the velocity and Q models. Because the Q model has not yet been updated, the dimming amplitudes in the seismic image have no improvements in Figure 11 and Figure 13.

During the final 20 iterations, I mainly focus on updating the Q model. Before Q updating, the seismic image (Figure 23) migrated using the updated velocity model and the initial Q model has only been compensated by the background Q value of 500. The image under the interpreted Q anomalies is wiped out. In my previous studies (Shen et al., 2013, 2014), I have shown that the parameter slope (ρ) is able to effectively quantify the attenuation effects from a seismic image. The slope value (ρ) can be computed from the logarithm of the spectral ratio between a measured spectra and a reference spectra (Tonn, 1991). The negative value of ρ means the image is undercompensated; while the positive value of ρ means the image is overcompensated. The larger the absolute value of ρ, the larger the Q effects measured from the seismic image, and the further the current Q model is from the accurate Q model – and vice versa. Therefore, the objective of this image-based Q inversion is to minimize the summation of $\rho(x; Q)$ over each image point. I have derived two objective functions to measure ρ from different image domains: (1) migrated image at zero subsurface offset (a postack objective function) and (2) ADCIGs (a prestack objective function). As described by Shen et al. (2013, 2014), inversion using a prestack objective function is able to produce a higher resolution and more accurate Q inversion results, although its computational cost is higher than the inversion using a poststack objective function.

To perform the inversion using a poststack objective function, I choose the trace at 33,880 m as the reference trace to compute ρ. This trace is far from both the shallow gas and the channel, and is assumed to be the least affected by the attenuation caused by these two Q anomalies. The window size to compute the spectra from the migrated image for ρ is 500 m in z direction and 125 m in x direction. I have developed an automatic picker to find the peak of the computed spectra over the wavenumber and record it as k_c. The desired wavenumber range that I choose to compute ρ is $[k_c - 0.015\ 1/\text{m},\ k_c + 0.015\ 1/\text{m}]$. The velocity stretch effect in the migrated image is also corrected before spectral analysis for ρ.

Figure 16(a) is the logarithm of the spectral ratio from the attenuated image (Figure 23) between the window below the left Q anomaly in which the window center is at $x = 27,500$ m, $z = 2,500$ m and the reference window in which the window center is at $x = 33,880$ m, $z = 2,500$ m. Figure 16(b) is the logarithm of the spectral ratio from the attenuated image (Figure 23) between the window far from these two Q anomalies in which the window center is at $x = 34,500$ m, $z = 2,500$ m and the reference window in which the window center is at $x = 33,880$ m, $z = 2,500$ m. The picked wavenumber range of these two locations is roughly within [0.005 1/m, 0.035 1/m]. Within this wavenumber range and regardless of the noisy wiggles, the curve in Figure 16(a) decreases linearly with wavenumber with a negative slope; while the curve in Figure 16(b) is almost flat with its slope value approximating to 0. These two slope values indicate the chosen window below the left Q anomaly is attenuated, and the chosen window far from these two Q anomalies has almost no attenuation. To remove the influence of the noise on the spectra shown as the wiggles, I use linear least-squares regression to fit a line to the curves of the logarithm of the spectral ratio within the picked wavenumber range by linear least squares regression. Figure 17 is the slope estimate of the image in Figure 23 for every image point used as the window center. The slope values at certain image points of the attenuated image in Figure 23 are positive. The possible reasons are because the chosen reference trace does not necessarily have the least attenuation impacts from the anomalies, and the way to pick the wavenumber range is not sophisticated enough. I clipped the positive numbers in Figure 17 to display the attenuated region only. The blue color in the slope map indicates the areas strongly attenuated. In accordance with the observation from Figure 23, two strong attenuation regions under the salt domes are highlighted by the blue in Figure 17. However, the blue regions below a 4 km depth are unexpectedly distributed, possibly because of the poor image quality below the base of salt. Fortunately, this deep part of the image has little impact on the estimation of the shallow Q anomalies.

Figure 18 is the inverted Q model displayed in logarithm scales ($\log_{10} Q$) after 20 iterations using the Q inversion with a poststack objective function. The results have two anomalies above the salt that are both shifted to the left of these two salt domes. The method using a poststack objective function has the disadvantage of producing a low-resolution result, which is part of the reason for the bias between an estimated location and the interpreted location of the anomalies. In addition, the right anomaly has a stronger attenuation than the left one, which contradicts the interpretation that the gas on the left has stronger attenuation than the right channel.

To perform the inversion using a prestack objective function (Shen et al., 2015), I choose the angle gather in which the midpoint is at 33,880 m as the reference gather to compute ρ, which is considered to be the least attenuated by two Q anomalies. This angle gather

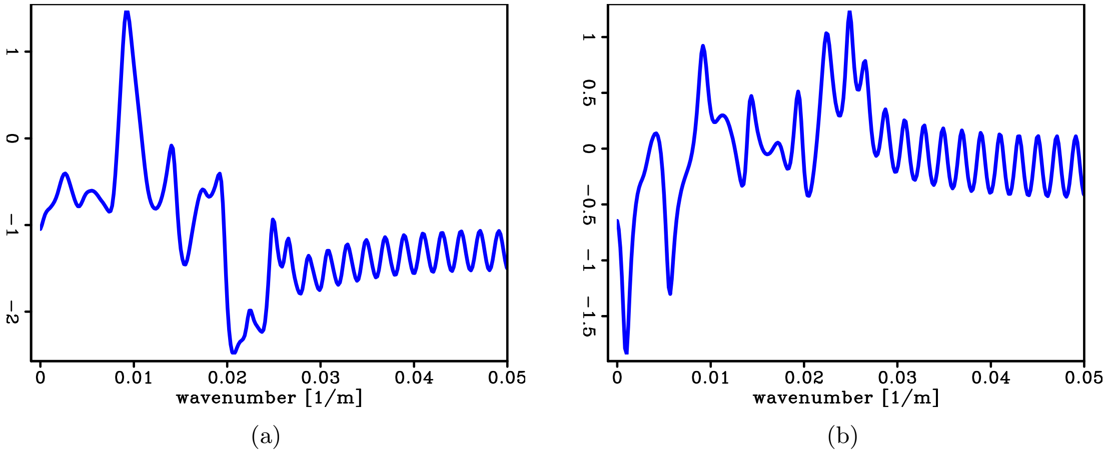


Figure 16: (a) The logarithm of the spectral ratio from the attenuated image (Figure 23) between the window below the left Q anomaly in which the window center is at $x = 27,500$ m, $z = 2,500$ m and the reference window in which the window center is at $x = 33,880$ m, $z = 2,500$ m. (b) The logarithm of the spectral ratio from the attenuated image (Figure 23) between the window far from these two Q anomalies in which the window center is at $x = 34,500$ m, $z = 2,500$ m and the reference window in which the window center is at $x = 33,880$ m, $z = 2,500$ m. **[CR]** yishen1/. dlp2D-iqtv-post-basc-rati-att,dlp2D-iqtv-post-basc-rati-ref

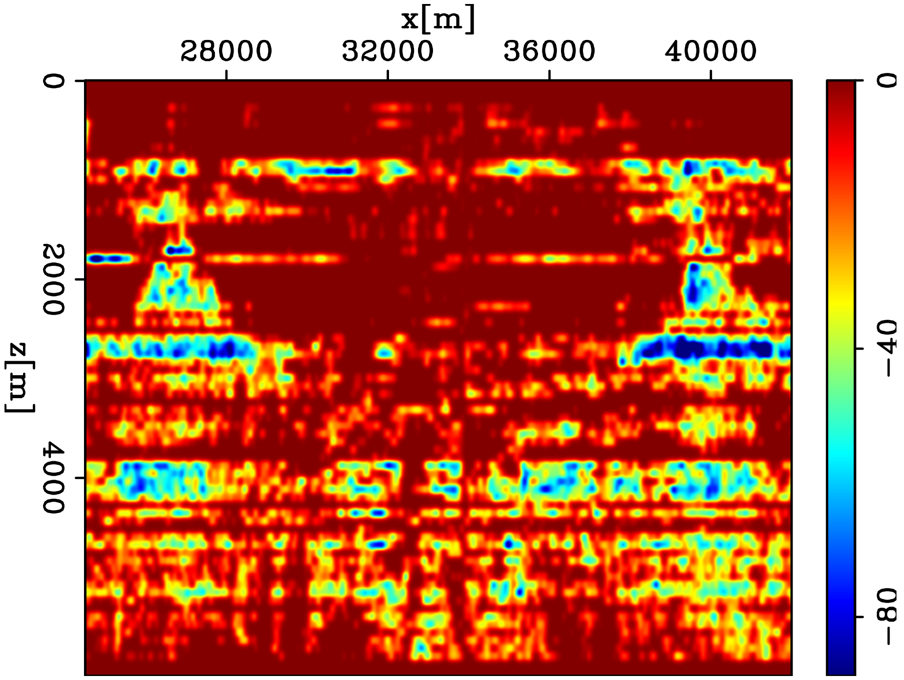


Figure 17: The slope estimate of the image in Figure 23 for every image point that is used as the window center. **[CR]** yishen1/. dlp2D-iqtv-post-basc-slopsft-iter0

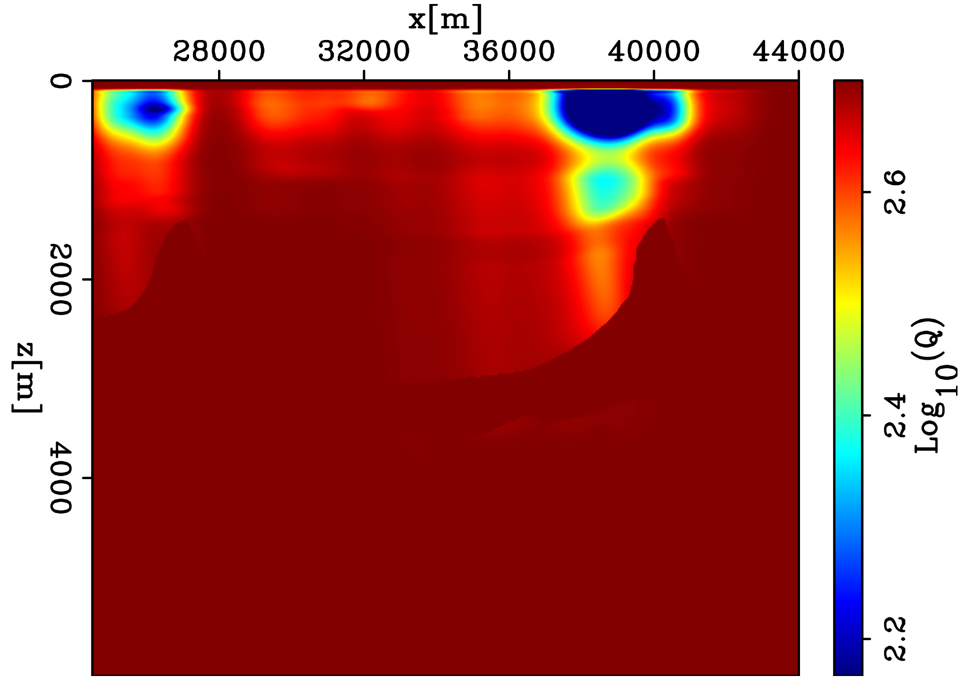


Figure 18: The inverted Q model displayed in logarithm scales ($\log_{10} Q$) after 20 iterations using the Q inversion with a poststack objective function. [**CR**] yishen1/. dlp2D-iqtv-post-basc-bq-iter20

is referenced to the rest of the angle gathers to compute the slope. The window size to compute the spectra from the migrated image for ρ has only one direction along the depth with the length of 300 m. The smaller window size aims for a higher-resolution Q inversion results. The automatic picker for the desired wavenumber range is the same as in the method using a poststack objective function. The velocity stretch effect and angle stretch effect (Shen et al., 2015) in the ADCIGs are also corrected before spectral analysis for ρ.

The ADCIGs at the first iteration of inversion mainly for Q model updating have only been compensated by the background Q value of 500. Figure 19 displays the slope estimate of these 9 representative ADCIGs. Every point in the angle gathers are the center of the windows for computing the slope . The slope values at certain points of the ADCIGs are positive because of the same reason previously described for the method using the poststack objective function. I clipped the positive number in Figure 19 to display the attenuated regions only. The blue color in the slope map indicates the strongly attenuated area. The blue colors are dominated at the depth below Q anomalies ($z < 1,500$ m) and above the base of the salt ($z > 4,000$ m). Again, the image below the base of the salt ($z > 4,000$ m) has little impact on the Q inversion. Therefore, I only focus the analysis below Q anomalies ($z < 1,500$ m) and above the base of salt ($z > 4,000$ m). With this depth range, the gathers in which the midpoint is far from the Q anomalies area (e.g., $x = 31$ km, $x = 33$ km and $x = 35$ km in Figure 28) have little attenuated region indicated by blue. As the midpoint moves closer to the Q anomalies (e.g., $x = 25$ km, $x = 29$ km and $x = 37$ km in Figure 28), the wave propagating at large angles passes through the Q anomalies; and therefore, the far angles at these ADCIGs are attenuated. Once the midpoint is at the anomalies area (e.g., $x = 27$ km, $x = 39$ km and $x = 41$ km in Figure 28), their near angles are attenuated. Figure 20 shows the zero-angle slope estimated measured from this attenuated ADCIGs with its blue color highlighting two strong attenuated regions within the salt domes.

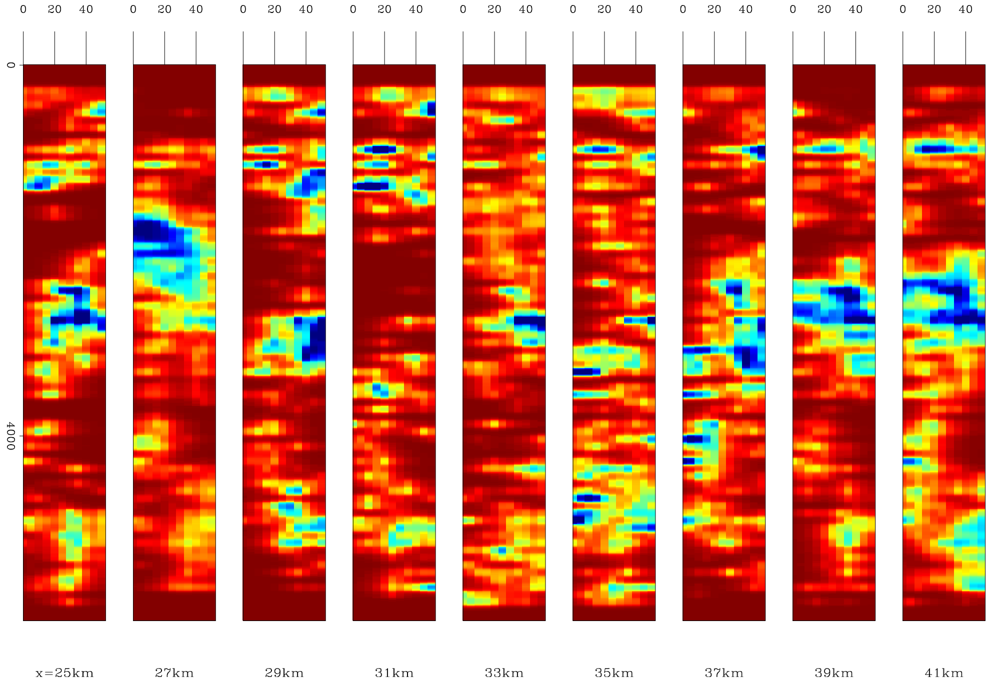


Figure 19: The slope estimate of the ADCIGs at the first iteration of inversion mainly for Q model updating. The slope estimates of 9 representative ADCIGs are displayed. Every point in the angle gathers are the center of the windows for computing the slope. [**CR**] yishen1/. dlp2D-iqtv-pre-basc-slopsft-iter0

Figure 21 shows the Q gradient at the first iteration of inversion mainly for Q model updating with a prestack objective function, in which the sign is opposite to the search direction. The gradient shows updates are strong around the shallow gas and channel region that exhibit the highest attenuation. Figure 22 is the inverted Q model displayed in logarithm scales ($\log_{10} Q$) after 20 iterations for Q inversion. The inverted Q models recover two Q anomalies. The location and the relative shapes of such Q anomalies in Figure 22 match the interpretation much better than the ones in Figure 18. This result also proves the conclusion that the inversion using a prestack objective function has a better capability to build an accurate Q model than the one using a poststack objective function.

Figure 24 is the migrated image at zero subsurface offset using the updated velocity model and the inverted Q model. The structures at the gas and channel regions become brighter. These bright spots matches the interpretation that the strong impedance contact makes seismic reflection amplitudes strong. The compensated image below the Q anomalies in Figure 24 are improved in terms of their enhanced amplitudes and higher frequency content. To take a closer examination, I zoom in the region around the left and right of the salt peaks in Figure 25 and Figure 26, respectively. Figure 25(a) and 26(a) are the zoomed-in regions of the attenuated image in Figure 23. Figure 25(b) and 26(b) are the zoomed-in regions of the compensated image in Figure 24. Because the enhancement of the absolute amplitudes by Q compensation is already obvious in Figure 24, I apply Automatic

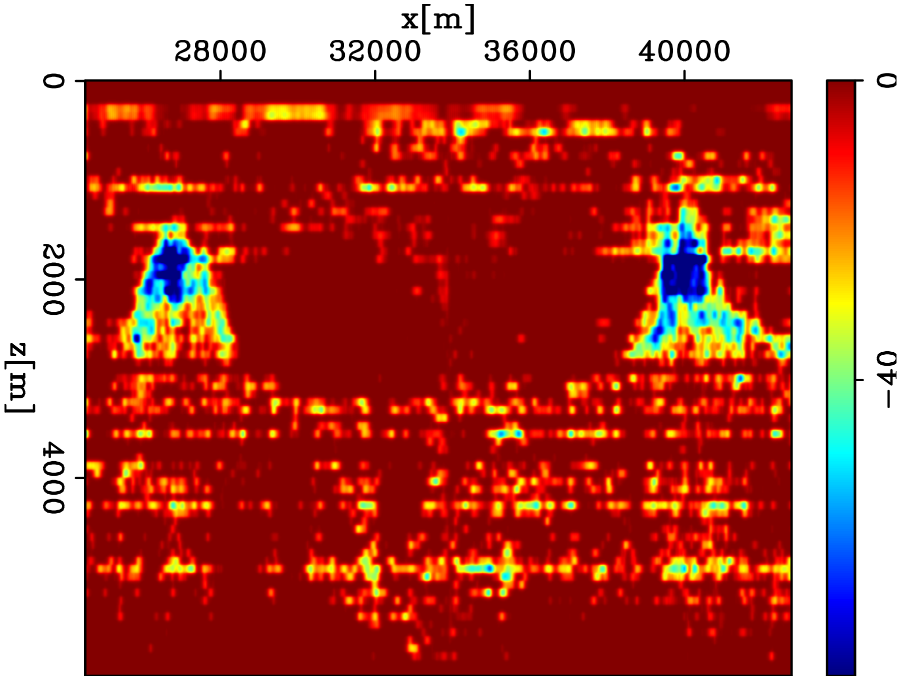


Figure 20: The slope estimate of the ADCIGs at the first iteration of inversion mainly for Q model updating. The slope estimates extracted at the zero angle are displayed. [**CR**] yishen1/. dlp2D-iqtv-pre-basc-slop0-iter0

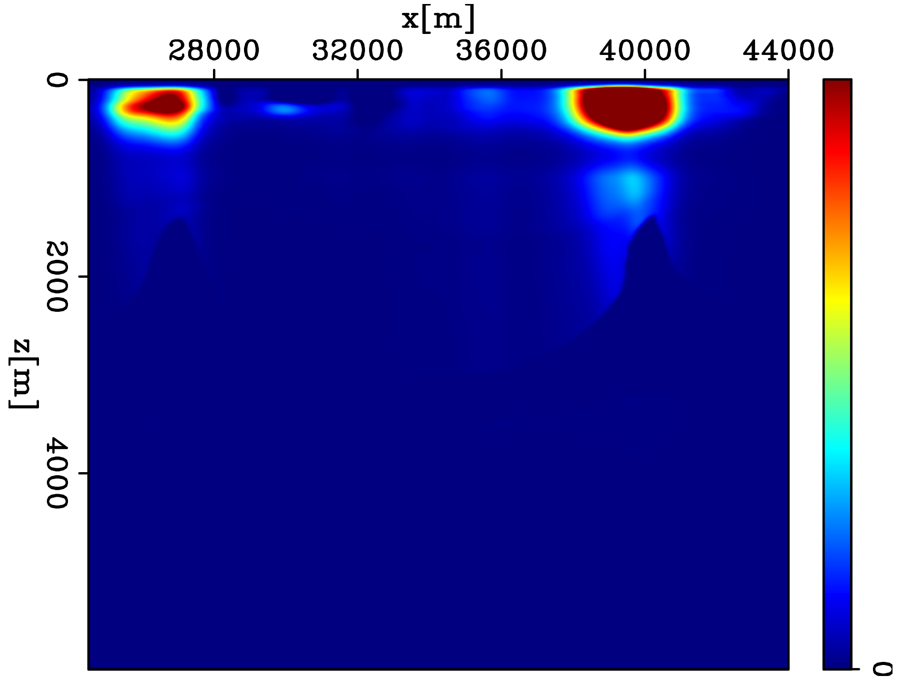


Figure 21: The Q gradient at the first iteration of inversion mainly for Q model updating, in which the sign is opposite to the search direction. [**CR**] yishen1/. dlp2D-iqtv-pre-basc-grdp-iter0

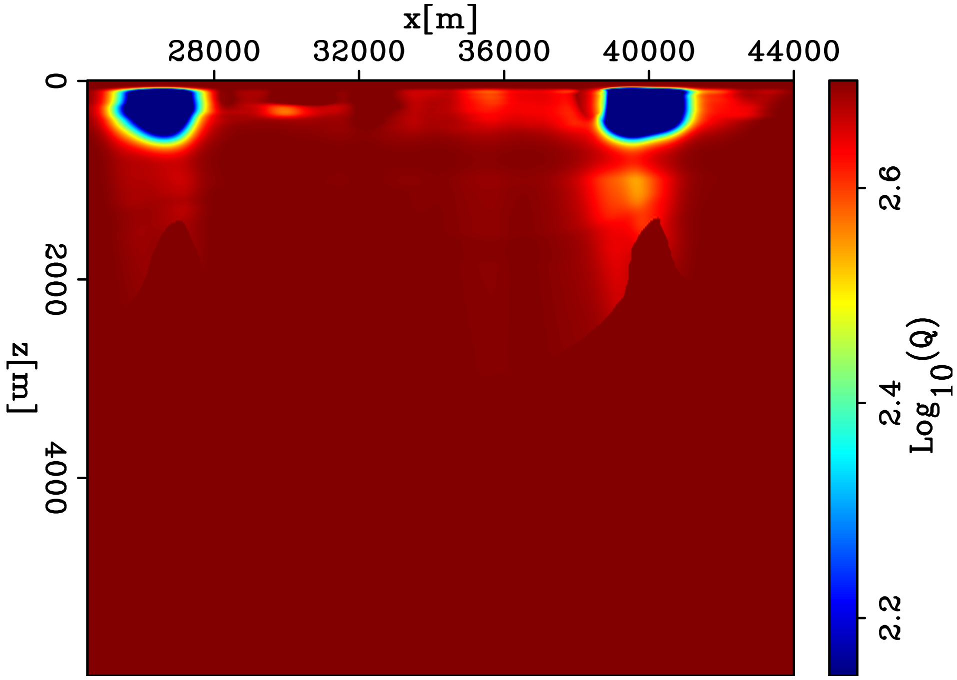


Figure 22: The inverted Q model displayed in logarithm scales ($\log_{10} Q$) after 20 iterations using the Q inversion with a prestack objective function. [**CR**] yishen1/. dlp2D-iqtv-pre-basc-bq-iter20

Gain Control (AGC) to the zoomed-in regions in Figure 25 and Figure 26 to compare their frequency and phase improvements. The compensated images show that events become sharper because of the recovery of high frequencies and become more coherent because of attenuation-induced phase distortion being corrected. I compute the spectra of the events below the right Q anomaly of Figure 23 and Figure 24, and display them in the logarithm scale in Figure 27. The spectra shows the compensation with the inverted Q model whitens the spectra and compensates for the higher frequency loss caused by attenuation.

Figures 28 and 29 show 10 representative ADCIGs of this migrated image with the Q model before and after Q updating, respectively. Results show the ADCIGs compensated by the inverted Q model have sharper and more coherent events. Such improvement brings benefit to the gather picking for velocity analysis and to AVO analysis for reservoir characterization.

CONCLUSION

The Dolphin's multi-client field data acquired in the North Sea used in this study has attenuation and velocity problems. Gas chimneys and channels exist in the subsurface with strong attenuation and low-interval velocity. In this paper, I first preprocessed this field data to regularize the bin grids and removed the swell noise, multiples, bubble and ghost. Then, I applied wave-equation migration velocity analysis to update the current velocity model. As a result, the velocity decreases in the gas and the channel area. The angle gathers migrated using the updated velocity model are much more flattened, and the events above

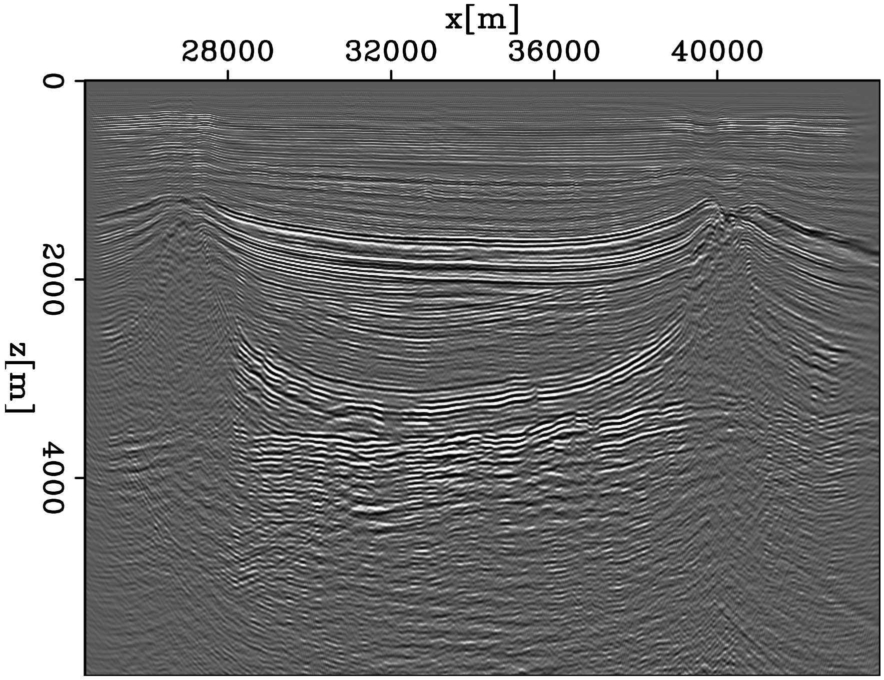


Figure 23: The migrated image at zero subsurface offset using the updated velocity model and the initial Q model. Same as Figure 11. Replicate the figure here for a convenient comparison with Figure 24. [**CR**] yishen1/. dlp2D-iqtv-pre-basc-bimg-before

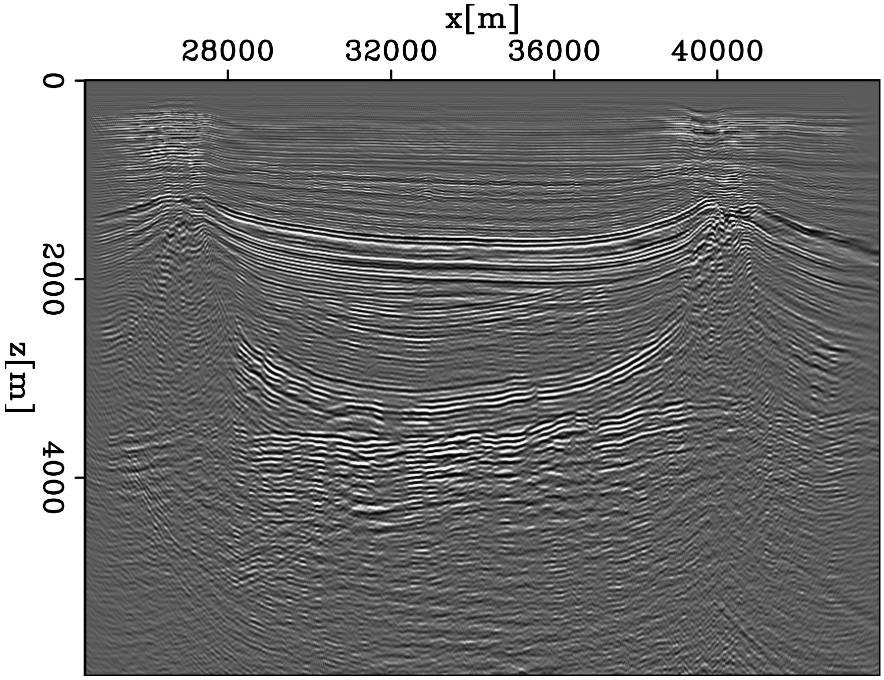
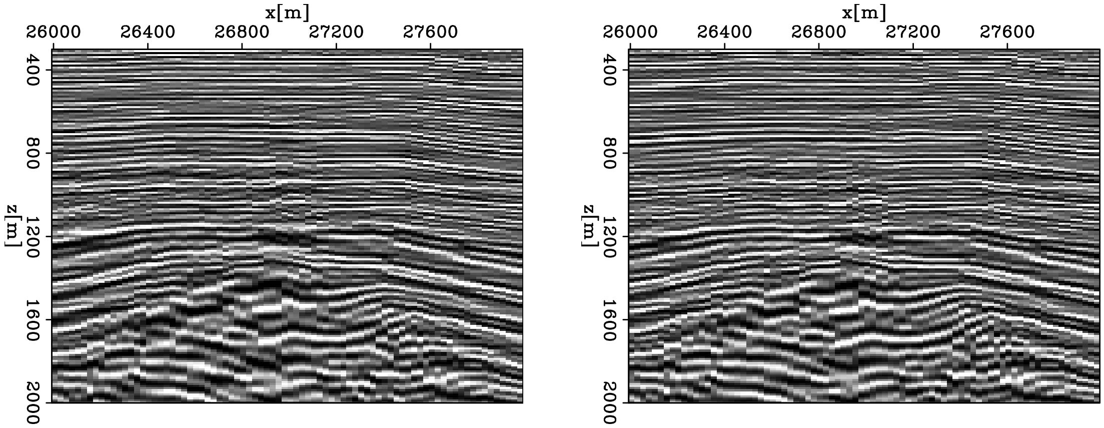
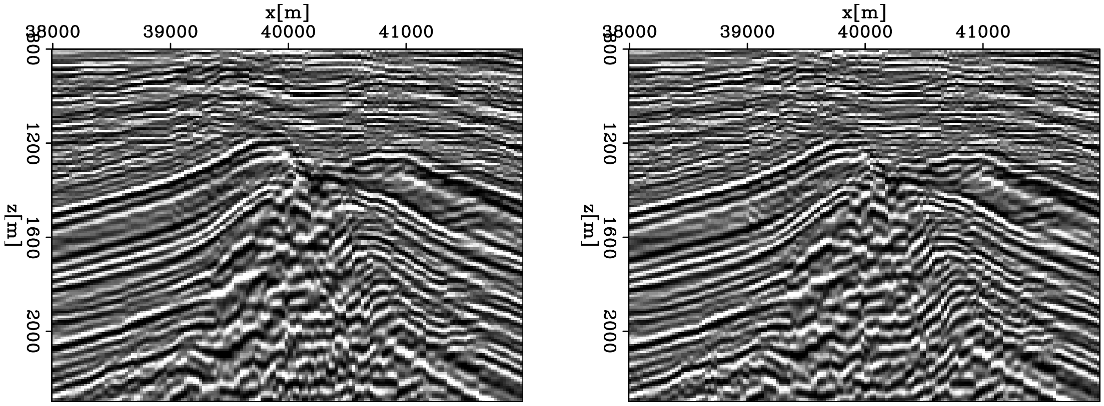


Figure 24: The migrated image at zero subsurface offset using the updated velocity model and the updated Q model. [**CR**] yishen1/. dlp2D-iqtv-pre-basc-bimg-after



(a) (b)

Figure 25: Zoomed-in region around the left of the salt peak of (a) attenuated image in Figure 23 and (b) compensated image in Figure 24. AGC is applied to both images. [**CR**] yishen1/. dlp2D-iqtv-pre-basc-reg1-before,dlp2D-iqtv-pre-basc-reg1-after



(a) (b)

Figure 26: Zoomed-in region around the right of the salt peak of (a) attenuated image in Figure 23 and (b) compensated image in Figure 24. AGC is applied to both images. [**CR**] yishen1/. dlp2D-iqtv-pre-basc-reg2-before,dlp2D-iqtv-pre-basc-reg2-after

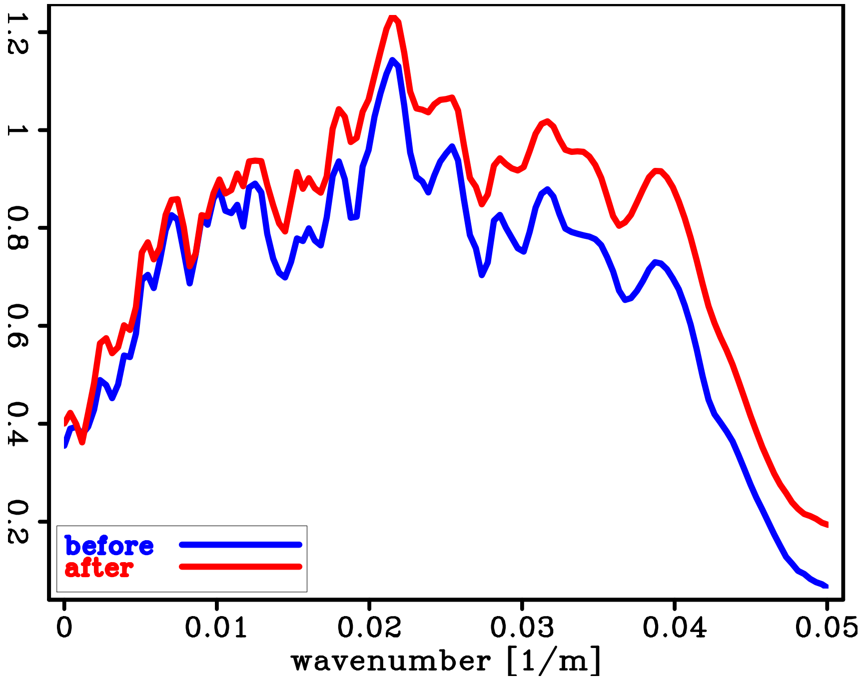


Figure 27: The spectra of the events below the right Q anomaly of Figure 23 and Figure 24. The spectra are displayed in the logarithm scale. [**CR**] yishen1/. dlp2D-iqtv-pre-basc-spex

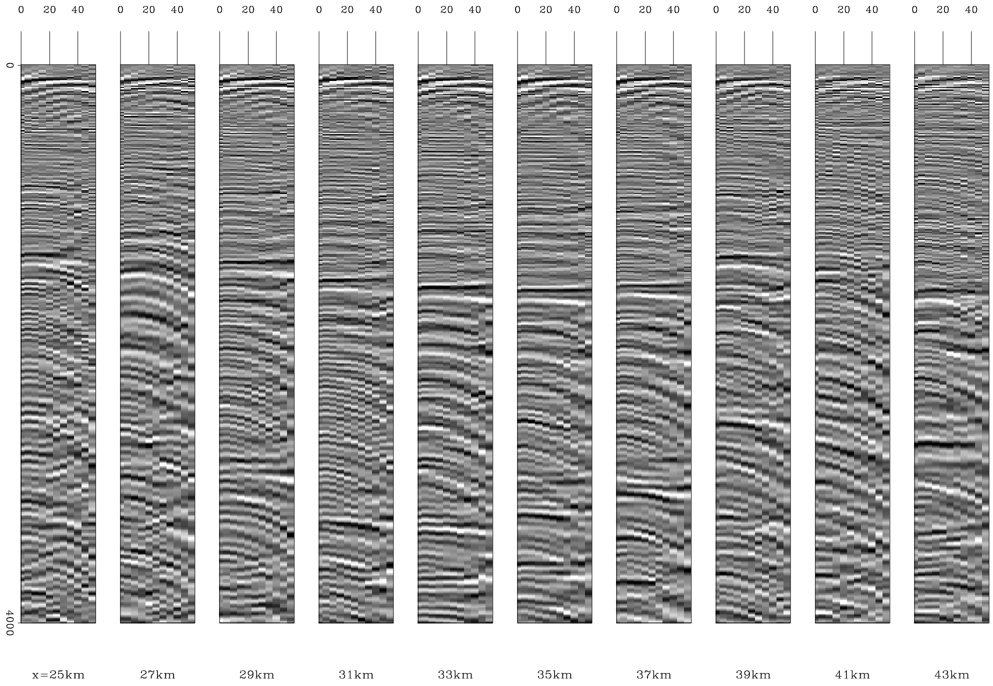


Figure 28: The ADCIGs using the updated velocity model and the initial Q model. AGC is applied. [**CR**] yishen1/. dlp2D-iqtv-pre-basc-bang-before

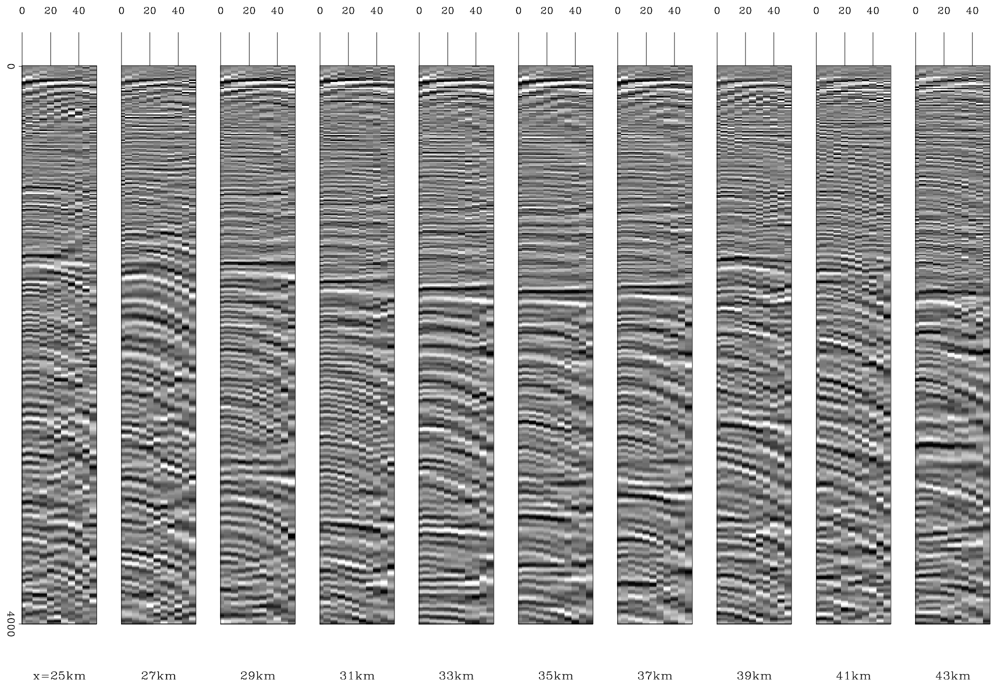


Figure 29: The ADCIGs using the updated velocity model and the inverted Q model. AGC is applied. [**CR**] yishen1/. dlp2D-iqtv-pre-basc-bang-after

the top of salt in the migrated images after velocity updating are more coherent. Then, I applied wave-equation migration Q analysis to invert for the Q models. The results show that the Q inversion using a prestack objective function better recovers the two Q anomalies than the Q inversion using a poststack objective function. The inverted Q model matches the interpretation quite well. Using this Q model in seismic migration, I made the seismic events below the anomalies clearly visible, with improved frequency content and coherency of the events.

ACKNOWLEDGMENTS

I would like to thank Dolphin Geophysical for providing these SHarp seismic data. I would also like to thank Biondo Biondi, Bob Clapp, Alejandro Cabrales, Gustavo Alves, Shuki Ronen, Stewart Levin, Jack Dvorkin, Ohad Barak, Taylor Dahlke for their help and discussions.

REFERENCES

Clarebout, J. and S. Fomel, 2014, Geophysical image estimation by example.

Sava, P. and B. Biondi, 2004, Wave-equation migration velocity analysis-i: Theory: Geophysical Prospecting, **52**, 593606.

Shen, Y., 2015, Simultaneous inversion of velocity and q using wave-equation migration analysis: SEP-Report, **160**, 71–80.

Shen, Y., B. Biondi, and R. Clapp, 2015, Wave-equation based Q tomography from angle-domain common image gathers, 4334–4338.

Shen, Y., B. Biondi, R. Clapp, and D. Nichols, 2013, Wave-equation migration Q analysis (WEMQA): EAGE Workshop on Seismic Attenuation Extended Abstract.

———, 2014, Wave-equation migration Q analysis (WEMQA): SEG Technical Program Expanded Abstracts.

Shen*, Y., C. Willacy, and V. Goh, 2015, Image-based Q tomography using wavefield continuation in the presence of strong attenuation anomalies: A case study in gulf of mexico.

Tonn, R., 1991, The determination of seismic quality factor Q from VSP data: A comparison of different computational techniques: Geophysical Prospecting, **45**, 87–109.

Zhang, Y. and B. Biondi, 2013, Moveout-based wave-equation migration velocity analysis: Geophysics, **78**, U31–U39.

Systematic bandwidth picking for the spectral ratio method

Stuart Farris

ABSTRACT

We use the spectral ratio method to estimate the effect of anelastic attenuation. The proper bandwidth over which to use the method is not constant throughout the entire image of the subsurface. Hand picking the correct bandwidth at every location would be prohibitive for a large 2D, let alone 3D image. Herein, various techniques are explored to systematically identify the upper and lower frequency bounds in order to improve performance of the spectral ratio method over large data sets. These techniques are evaluated with a visual metric to asses the accuracy of their bandwidth picks.

INTRODUCTION

Anelastic attenuation is represented by the Q factor in reflection seismology. A good estimate of the Q factor is useful for reflection data processing and rock property prediction. For example, processing methods, such as migration, Q filtering, and deconvolution that account for attenuation can produce higher quality images than methods which ignore it. Further, attenuation estimates can reveal information about porosity, permeability, and fluid saturation useful for reservoir characterization (Best et al., 1994).

There are numerous techniques that attempt to accurately predict the Q factor within a seismic image. The spectral ratio method is regarded as one of the more robust techniques (Tonn, 1991). A benefit of the spectral ratio method is that it can be performed over a limited range of frequencies that are signal dominated, hence disregarding noise or artifact dominated parts of the spectrum. While this makes the technique more versatile, it also implies the signal dominated bandwidth of the spectra must be chosen beforehand. Moreover, the proper bandwidth is not constant throughout the seismic image, implying it must be independently chosen at every location. Manually picking this proper bandwidth at every location would be extremely time consuming.

If a systematic technique could be developed to accurately pick the signal dominated bandwidth, the speed at which the spectral ratio method could be performed over an entire seismic image would increase dramatically. Consequently, the speed at which the Q factor, the output of the spectral ratio method, could be found at every point in a model would also increase. The focus of this study is picking the proper bandwidth for the spectral ration method.

I considered many techniques to pick the correct bandwidth for the spectral ratio method. It was logical to begin with the simplest methods and progress towards more complex ones. Those tested here:

1. use a constant, narrow bandwidth that might represent signal dominated frequencies at all locations,
2. fix the lower frequency bound and linearly decrease the upper bound with depth,
3. center the bandwidth around the peak frequency.

I also discuss other techniques that may perform better than those tested in this report.

SPECTRAL RATIO METHOD

The spectral ratio method can be used to determine the Q effect over some discretization of a seismic image by comparing incident and outgoing waves. Take some outgoing wave, $R(f)$, which is the result of some incident wave, $S(f)$, and the attenuating medium response $GH(f)$:

$$R(f) = GH(f)S(f) \tag{1}$$

where f is frequency and G represents the geometrical spreading, instrument response, source and receiver coupling, radiation pattern, and reflection/transmission coefficients. $H(f)$ includes the attenuation effect on amplitude and was shown by Ward and Toksz (1971) to be expandable to

$$H(f) = \mathrm{e}^{-f \int_{ray} \frac{\pi}{Qv}}, \tag{2}$$

where v is the wave velocity. Substituting $H(f)$ into equation 1 gives

$$R(f) = G\mathrm{e}^{-f \int_{ray} \frac{\pi}{Qv}} S(f). \tag{3}$$

With some manipulation, equation 3 becomes:

$$ln[R(f)/S(f)] = (\int_{ray} \frac{\pi}{Qv})f - ln(G). \tag{4}$$

This reveals that the relation between $R(f)$ and $S(f)$ is linear with respect to frequency following

$$y(f) = mf + b, \tag{5}$$

where

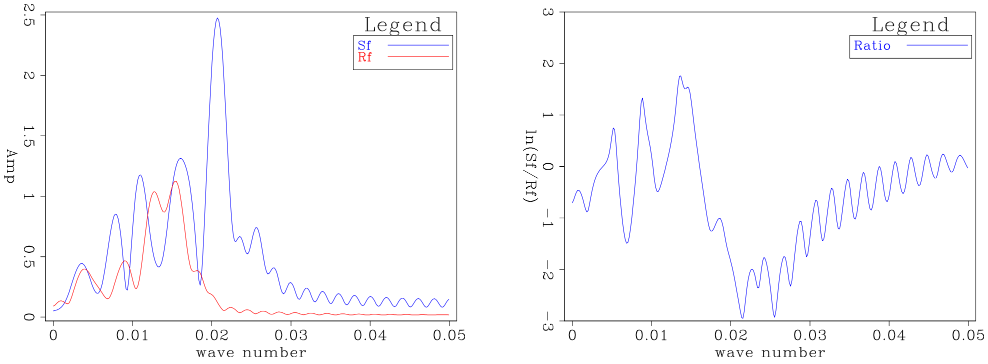
$$y(f) = ln[R(f)/S(f)], \tag{6}$$

$$m = (\int_{ray} \frac{\pi}{Qv}), \tag{7}$$

$$b = -ln(G). \tag{8}$$

Therefore, Q can be found using the slope of the line representing the log of the ratio between the input spectrum, $S(f)$, and the output spectrum, $R(f)$. Notice that the b term can be ignored which means this method is not affected by G and its dependencies.

In practice, the seismic image is discretized into spectrum windows. Each window is of the order of tens of meters wide and hundreds of meters tall. Windows overlap to ensure spectra are not dominated by local reflectivity. To calculate Q at each window, the spectrum is treated as the outgoing $R(f)$ and is divided by some input spectrum $S(f)$. The input spectrum $S(f)$ does not necessarily need to be the source spectrum, but can be some spectrum from a non-attenuated window. I will call this the reference window. The log of this ratio is calculated and the line of best fit is found. Using Equation 7, Q can be found from the slope of this line. Figure 1 illustrates an example of $R(f)$, $S(f)$, and their log ratio.



(a) (b)

Figure 1: a) Spectra of some window of interest and a reference window, b) log ratio of the reference spectra divided by the window of interest. [**ER**] sfarris1/. spexref-mx600-win30,rati-mx600-win30

An added advantage to the spectral ratio method is that is can be applied over a limited range of frequencies to avoid those dominated by noise. As seen in Figure 1, data above wavenumber 0.03 represent noise and rather than signal. If linear regression is performed over all wave numbers, the resulting line and its slope will be largely affected by the noise dominated region. The noise can be avoided by evaluating the line only over the signal dominated wavenumbers. Picking the correct frequencies to evaluate is the purpose of this investigation.

DATA

The data used in this analysis is Dolphin Geophysical's multi-client field data acquired in the North Sea. There are significant velocity and attenuation challenges caused by a salt body with two peaks. As interpreted by Dolphin, a gas chimney is present above one of the salt peaks and a channel above the other. The gas chimney forms a migration pathway for gas to leak and accumulate at shallow depths above the salt. The shallow gas gives rise to strong attenuation and low interval velocity. The channel also has low velocities and provides strong attenuation. These regions of strong attenuation, and therefore significant Q effects, make the dataset ideal for testing the spectral ratio method.

As previously mentioned, the seismic image will be discretized into spectra windows. The window size used in this study is 500 m in the z direction and 125 m in the x direction. The spectrum at each window will be the total spectrum from a window of this size centered at the window's location in the subsurface.

BANDWIDTH PICKING TECHNIQUES

Herein I present the various methods tested for choosing the correct bandwidth over which to perform the spectral ratio method.

Constant bandwidth

The most simple technique is to use a constant bandwidth for all windows. A bandwidth should be chosen that could reasonably represent the frequencies dominated by signal for all depths and locations in the subsurface.

Linearly shrinking bandwidth

With this technique the lower frequency bound is fixed and the upper bound decreases linearly with depth. Since higher frequencies are expected to naturally attenuate with depth, it follows that the bandwidth dominated by signal should also shrink with depth. I implement two variations of this method. In the first, I fix the lower bound at wavenumber 0.0 and linearly move the upper bound from wavenumber 0.04 to 0.02. In the second, I fix the lower bound at 0.005 and linearly move the upper bound from wavenumber 0.04 to 0.02. I vary the lower bound in an attempt to avoid low frequency noise.

Center around peak

The frequency with the largest amplitude most likely represents the location with the strongest signal. By centering the bandwidth around this region, it is reasonable to assume it represents a bandwidth dominated by signal. This bandwidth should be wide enough to capture enough frequencies to produce a spectral method result representative of the signal dominated region, but narrow enough to avoid noise dominated frequencies. Here, I used a width of 0.015 wavenumbers centered around the peak frequency.

RESULTS

A metric is required to compare the results of each picking method. I use a visual metric focusing on the slope of the log ratio used in the spectral ratio method.

At nine locations over a 2D slice of the Dolphin dataset, presented in Figure 2, the bandwidth dominated by signal was chosen manually. The slope was found from the log ratio of the picked window and a reference window. This will be the control slope. We can see that the windows at the center of Figure 2 are non-attenuated. These were used as the reference windows. At the nine locations, each picking method selected a particular bandwidth. The slope was then found over the bandwidth picked by each method. By visually comparing the manually picked slopes with the slopes picked by each method, the accuracy of each technique can be evaluated.

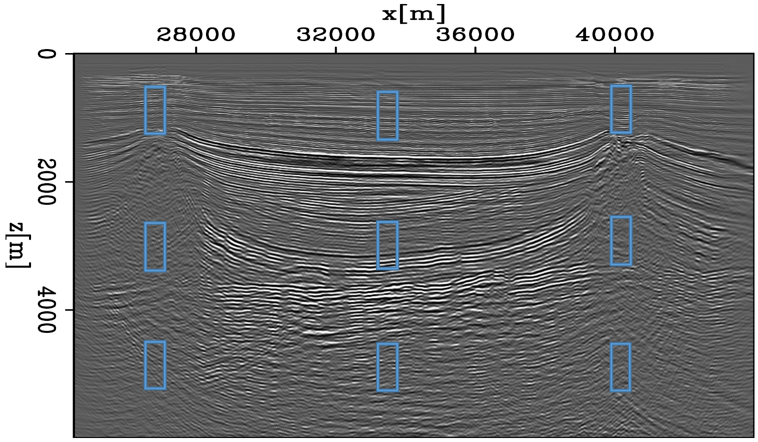
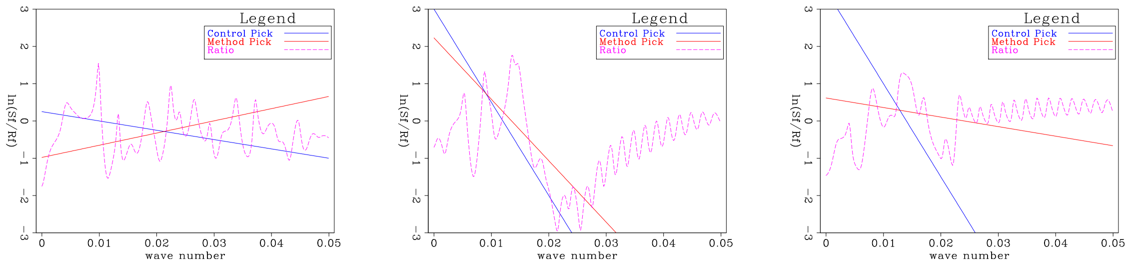


Figure 2: Location of nine visual metric locations. Size of windows not representative of actual window size. [**NR**] sfarris1/. metricWindows

Slope results

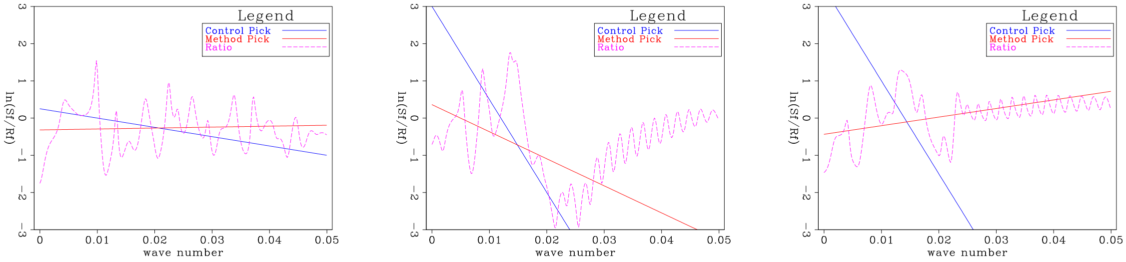
To illustrate the varying slope results for each technique, I chose three windows to analyze. The windows are located at 40,000 meters, as seen in Figure 2, with depths of 1000, 3000, and 5000 meters. These three windows were chosen because they represent regions in the subsurface with varying amounts of noise. Figures 3, 4, 5, and 6 show results for each of these windows. In each graph, there are three lines. The dashed, pink line is the log of the ratio between the reference window and the current window. Again, the slopes are calculated by linear regression over some bandwidth of this log ratio line. The solid blue line is the control slope pick. The solid red line is the slope picked by one of the designed techniques.



(a) (b) (c)

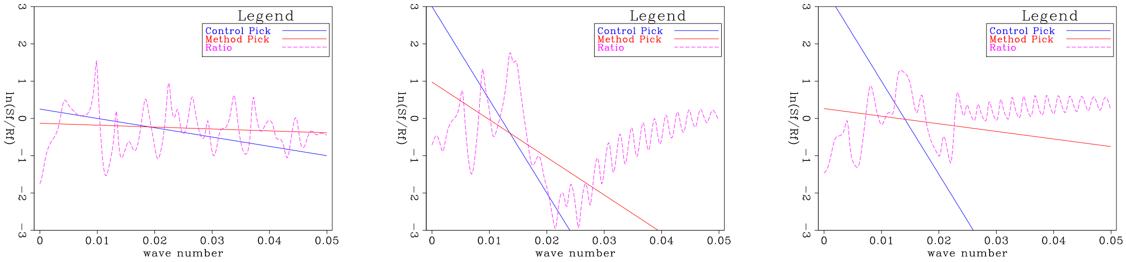
Figure 3: Log ratio (dashed pink), control slope pick (solid blue), and constant bandwidth method slope pick (solid red) at x=40000 m and, (a) z=1000 m, (b) z=3000 m, (c) z=5000 m [**ER**] sfarris1/. mx600-win10-b4,mx600-win30-b4,mx600-win60-b4

While the metric comparing slope picks is useful for analyzing accuracy at individual windows, it does not reveal how a technique performs throughout the entire image. To gain insight into the big picture results, the slopes were calculated at every window using each technique. These slope panels are displayed in Figure 7. A clip of 150 was applied to each panel.



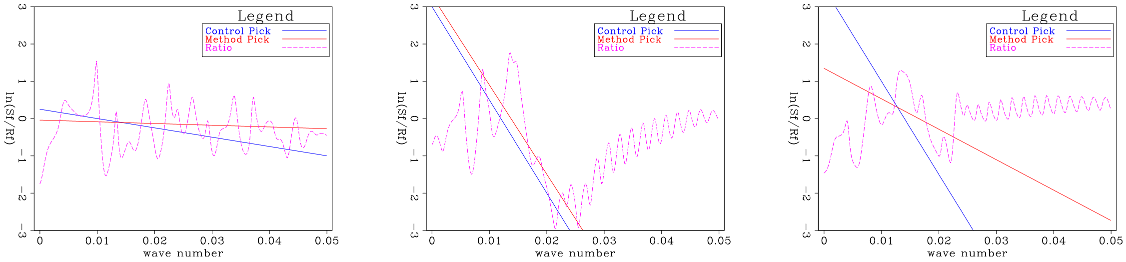
(a) (b) (c)

Figure 4: Log ratio (dashed pink), control slope pick (solid blue), and linearly shrinking bandwidth (min=wavenumber 0.00) slope pick (solid red) at x=40000 m and, (a) z=1000 m, (b) z=3000 m, (c) z=5000 m [**ER**] sfarris1/. mx600-win10-b1,mx600-win30-b1,mx600-win60-b1



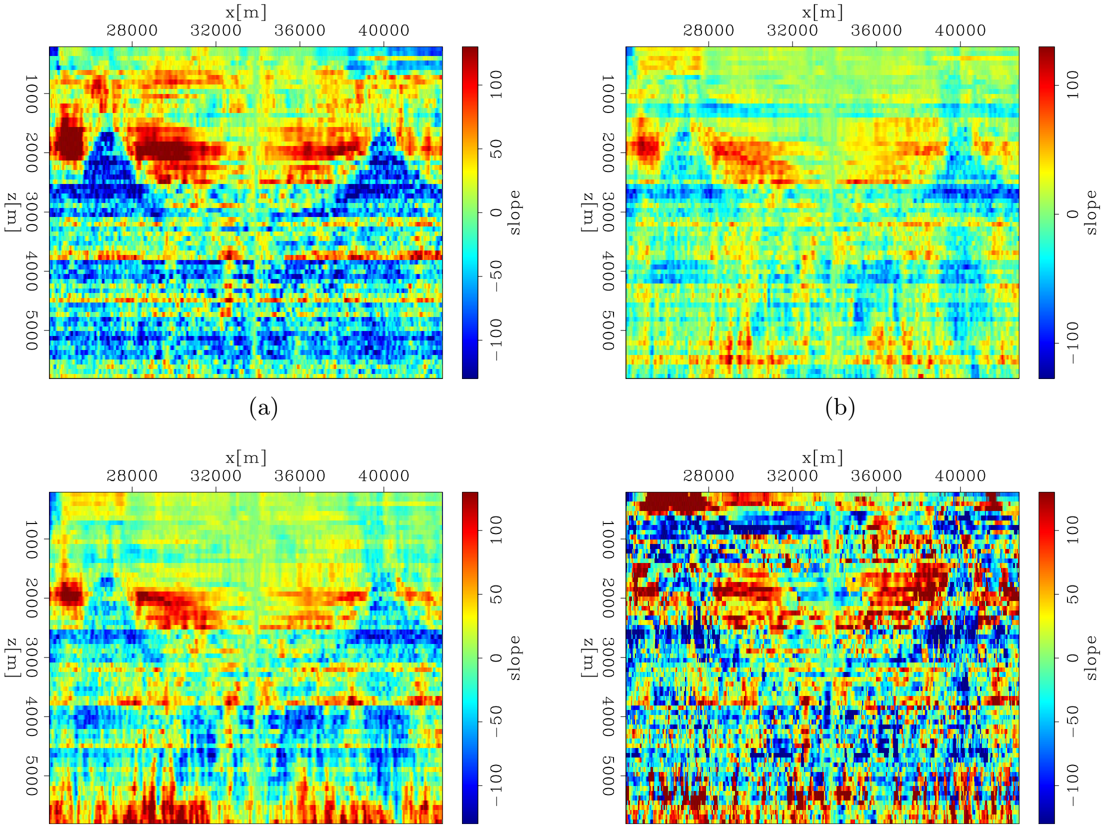
(a) (b) (c)

Figure 5: Log ratio (dashed pink), control slope pick (solid blue), and linearly shrinking bandwidth (min=wavenumber 0.005) slope pick (solid red) at x=40000 m and, (a) z=1000 m, (b) z=3000 m, (c) z=5000 m [**ER**] sfarris1/. mx600-win10-b3,mx600-win30-b3,mx600-win60-b3



(a) (b) (c)

Figure 6: Log ratio (dashed pink), control slope pick (solid blue), and bandwidth centered around peak slope pick at x=40000 m and, (a) z=1000 m, (b) z=3000 m, (c) z=5000 m [**ER**] sfarris1/. mx600-win10-b2,mx600-win30-b2,mx600-win60-b2



(a) (b)

(c) (d)

Figure 7: Slope values at every window for (a) constant bandwidth method, (b) linearly shrinking bandwidth (min= wavenumber 0.0), (c) linearly shrinking bandwidth (min= wavenumber 0.005), (d) bandwidth centering around peak frequency [**ER**] sfarris1/. slopePlot4,slopePlot1,slopePlot3,slopePlot2

DISCUSSION

Figure 6 shows the technique centered around the peak frequency best matched the control picks. This is a reasonable result since it is the only method that actually considers the spectrum of the current window. Other methods blindly choose bandwidths without dynamically considering where the signal is most dominant. Still, each method had trouble matching the control at large depths and often at shallow depths. As currently implemented, none of the methods can be considered accurate enough fot systematically picking the correct bandwidth over which to perform the spectral ratio method.

While the full slope panels in Figure 7 were not the main metric used to determine viability, they do provide unique insight. Primarily, these panels reveal a promising proof of concept. Each contains a slope trend that outlines the salt bodies beneath the shallow Q anomalies seen in the original data image. These slope trends almost appear as lenses beneath the Q anomalies. Since we would expect higher attenuation under these anomalies, it is promising that the techniques are finding slope trends that are directly related to lower Q values. Furthermore, the large positive slopes seen near these lenses are caused by the large size of each window. Indeed, the spectrum at each window is actually an average of a few traces 500 meters in length. This implies that some windows interact with regions affected by the shallow Q anomalies when they are actually located outside the lensing region. This interaction causes an anomalous log ratio result leading to a positive slope outcome. Therefore, these large positive anomalies are not caused by the bandwidth picking methods but by the large window size.

FUTURE WORK WITH NEW TECHNIQUES

A numeric metric should be created to determine the viability of each bandwidth picking method. One possibility is to find the mean square error (MSE) between the control slopes and the slopes found by each technique. This would quantify how far the results are from the control.

A more sophisticated bandwidth picking technique could be developed using machine learning. I can see two possible approaches. The first would use regression methods based on control bandwidth picks to find parametric estimations of the lower and upper frequency values. The second would be a clustering method analyzing various parameters such as signal-to-noise, amplitude, average frequency, or spikiness.

Finally, the window size used to discretize the image's spectrum could be altered. There exists an optimal size that eliminates local reflectivity but avoids issues on the borders of large Q anomalies.

ACKNOWLEDGMENTS

I would like to acknowledge Dolphin Geophysical for providing the SEP with the data used in this report.

A special thanks to Fantine Huot for introducing me to this project and mentoring me through my first SEP report. Further, Joseph Jennings for solving endless Unix, C++, and

Fortran issues. Of course, Yi Shen for allowing me to use the Dolphin data intended for her thesis.

Finally, the sponsors of the SEP should be recognized for their financial support.

REFERENCES

Best, A., C. McCann, and J. Sothcott, 1994, The relationships between the velocities, attenuations and petrophysical properties of reservoir sedimentary rocks1: Geophysical Prospecting, **42**, 151–178.

Tonn, R., 1991, The Determination of the Seismic Quality Factor Q from Vsp Data: A Comparison of Different Computational Methods1: Geophysical Prospecting, **39**, 1–27.

Ward, R. W. and M. N. Toksz, 1971, Causes of regional variation of magnitudes: Bulletin of the Seismological Society of America, **61**, 649–670.

Linearized Waveform Inversion with Velocity Updating: Theory and first results

Alejandro Cabrales-Vargas, Biondo Biondi, and Robert Clapp

ABSTRACT

Linearized Waveform Inversion or least-squares migration is a process that aims at obtaining a better estimation of the subsurface reflectivity, in comparison with conventional migration. During the process, the background model (velocity or slowness) remains invariant. Only the reflectivity component is updated.

In this report we revisit the Linearized Waveform Inversion with Velocity Updating theory introduced in a previous report, and present the first synthetic examples. The method introduces controlled updates to the background model during the reflectivity inversion, correcting for slowness inaccuracies that negatively affect seismic amplitudes during conventional linearized waveform inversion. The method incorporates Wave-Equation Migration Velocity Analysis to transform such background model updates into perturbations in the image.

INTRODUCTION

Imaging complex oil & gas reservoirs with seismic methods demands solutions beyond kinematically accurate subsurface images. Beyond identification of reservoir traps, proper characterization of oil-bearing rock facies is vital for optimal drilling programs and exploitation. One important tool in the achievement of these goals is the variation of seismic amplitudes as a function of the rock and fluid type.

For several years, the interpretation of amplitude variations has relied on "true-amplitude" Kirchhoff migration images. The high-frequency approximation assumption entailed by Kirchhoff modeling and migration algorithms allows splitting the solution into two components: a kinematic component, and an amplitude component. The former is obtained by means of traveltime computations, resulting in an image with correctly positioned seismic events in relatively simple geology. The latter (amplitude coefficients) is usually an approximation to the transport equation solution.

Kirchhoff-based solutions present an important caveat: the high-frequency approximation fails in the presence of strong velocity contrasts and complex geology. Downward continuation methods and two-way wave equation methods are thereby mandatory for adequately addressing such complexities. Nonetheless, the final image usually is only kinematically correct. One reason is that the imaging condition constitutes a zero-lag crosscorrelation of the source wavefield and the receiver wavefield (Claerbout, 1971). The amplitudes can be balanced by implementing deconvolution imaging conditions (e.g. Guitton et al., 2007), but the amplitude still remains inaccurate.

One important limitation of the migration methods, regardless of the amplitude treatment, is that they merely constitute the first step of the inversion process aimed at estimating the subsurface reflectivity:

$$\tilde{\mathbf{m}} = \mathbf{L}'\mathbf{d}_0, \tag{1}$$

where $\mathbf{d}_0$ represents the seismic data collected at discrete surface locations, $\tilde{\mathbf{m}}$ represents here the estimated reflectivity, and $\mathbf{L}'$ represents the *adjoint* of the Born modeling operator, better known as *migration* (Claerbout, 1992). We denominate $\tilde{\mathbf{m}}$ as the *migrated image*. Born modeling constitutes a linearization of the full modeling operator, $\mathcal{L}$, around a background velocity (or slowness) field, with the reflectivity acting as a perturbation in such background field.

Unfortunately, the amplitudes of the migration image are often not representative of the geologic variations. The problem is further exacerbated by irregular and/or sparse acquisition geometries (frequently due to obstacles and/or limited budget) and limited frequency bandwidth of the data. As a consequence, the resulting image often constitutes a blurred version of the subsurface reflectivity: the imperfect acquisition geometry and limited frequency content negatively impact both the seismic resolution and amplitude preservation.

Linearized waveform inversion (LWI) (*a.k.a.* least-squares migration) (e.g. Nemeth and Schuster, 1999; Ronen and Liner, 2000; Clapp, 2005; Valenciano, 2008; Tang, 2011b; Dai et al., 2010; Fletcher et al., 2016) constitutes an iterative process for approximating the inverse of seismic modeling, hence recovering a better reflectivity estimation than conventional migration. It can be expressed as the solution of the normal equations, formally cast as

$$\mathbf{L}'\mathbf{L}\mathbf{m} = \mathbf{L}'\mathbf{d}_0, \tag{2}$$

where $\mathbf{m}$ constitutes the subsurface reflectivity. Solving Equations 2 is equivalent to minimizing the following objective function in data space

$$\Phi(\mathbf{m}) = \frac{1}{2}\|\mathbf{L}\mathbf{m} - \mathbf{d}_0\|_2^2, \tag{3}$$

or in model space

$$\Phi(\mathbf{m}) = \frac{1}{2}\|\mathbf{H}\mathbf{m} - \tilde{\mathbf{m}}\|_2^2, \tag{4}$$

where $\mathbf{d} = \mathbf{L}\mathbf{m}$ constitutes synthetic data, and $\mathbf{H}$ is the *Hessian* operator associated to the misfit function Φ (*a.k.a. Gauss Newton Hessian*). Note that in Equation 3 we aim at fitting synthetic data to the recorded data, whereas in Equation 4 we aim at fitting a reflectivity model to the migrated image. Both constitute different versions of the same optimization problem, which is *linear* because the Born operator and the Hessian are linear with respect to the reflectivity model. In other words, such operators are *independent* of the reflectivity model.

LWI has been so far parameterized exclusively in terms of the subsurface reflectivity. The velocity field is assumed to be either correct or inaccurate to some degree (e.g. Luo and Hale, 2014), but either way fixed during the optimization process. It does not participate in the reflectivity estimation during LWI. Thus, the obvious question is, can we push the LWI capabilities further by allowing velocity (or slowness) to participate in the optimization process?

We developed Linearized waveform inversion with velocity updating (LWIVU) to incorporate a perturbation component of the velocity field into the optimization process

(Cabrales-Vargas et al., 2016). The objective is to obtain better estimations of the reflectivity with respect to conventional LWI. The updates in the velocity (or slowness) field are intended to be perturbations small enough to maintain the linearity of the inversion. Moreover, such perturbations are not incorporated back to the background velocity field, but merely used to compute perturbations in the image that would increase the accuracy of the inverted reflectivity. With this approach, we exploit the fact that, in real data, the "correct velocity field" is accurate enough for positioning seismic events, but remaining inaccuracies can affect the amplitudes.

LWIVU can be derived from a simplification of the nonlinear optimization process known as *full-waveform inversion* (FWI), by splitting the Hessian into Gauss-Newton Hessian and *wave-equation migration velocity analysis* (WEMVA) Hessian, and adding a WEMVA-based constraint that maximizes the stacking power. In the next section we revisit the original derivation introduced in our previous report (Cabrales-Vargas et al., 2016). Next, we present synthetic examples in a two-layer model with a Gaussian slowness anomaly. Next, we analyse the computational costs of LWIVU compared to conventional LWI. Finally, we present the conclusions of this report.

THEORY

Derivation of LWIVU from FWI

Full-waveform inversion (FWI) (Tarantola, 1984; Virieux and Operto, 2009; Fichtner, 2011; Biondi and Almomin, 2014) is a nonlinear optimization scheme that minimizes the misfit between the recorded seismic data, $\mathbf{d}_r$, and modeled data, $\mathbf{d}$, with respect to a subsurface parameters model, $\mathbf{m}$,

$$\Phi_{FWI}(\mathbf{m}) = \frac{1}{2}\|\mathbf{d} - \mathbf{d}_r\|_2^2. \tag{5}$$

where $\mathbf{d} = \mathcal{L}(\mathbf{m})$. As aforementioned, $\mathcal{L}$ constitutes the *non-linear seismic modeling operator*. It is worth remarking that the non-linearity of this operator is with respect to the model parameters, $\mathbf{m}$. On the contrary, the operator is linear with respect to the source term (Almomin, 2013).

We can apply the Gauss-Newton method to minimizing the misfit function in Equation 5. Firstly, notice that a necessary condition is that the gradient of the misfit function vanishes at a minimum (Fichtner, 2011),

$$\nabla\Phi_{FWI}(\mathbf{m}_{min}) = 0, \tag{6}$$

where $\mathbf{m}_{min}$ constitutes such a minumum. Now we can expand the gradient in Equation 6 in Taylor series around a nearby model, $\mathbf{m}_i$, such that the difference $\delta\mathbf{m} = \mathbf{m}_{min} - \mathbf{m}_i$ is small enough to allowing us to truncate the series at the second term:

$$\nabla\Phi_{FWI}(\mathbf{m}_{min}) \approx \nabla\Phi_{FWI}(\mathbf{m}_i) + \mathbf{H}(\mathbf{m}_i)\delta\mathbf{m} = 0, \tag{7}$$

where $\mathbf{H}$ represents the FWI Hessian operator. It is the second derivative of the misfit function with respect to the model parameters. We use this result to estimate the model update, $\delta\mathbf{m}$, by solving the linear system

$$\mathbf{H}(\mathbf{m}_i)\delta\mathbf{m} = -\nabla\Phi_{FWI}(\mathbf{m}_i), \tag{8}$$

which is known as the *Newton's equation.* In the FWI problem we solve Equation 8 as part of an iterative non-linear optimization scheme aimed at updating the model parameters using $\mathbf{m}_{i+1} = \mathbf{m}_i + \delta\mathbf{m}$, repeating the process until desired convergence is reached.

For the LWIVU derivation, we adopt the Barnier and Almomin (2014) notation, separating the model parameters in a low-wavenumber component, $\mathbf{b}$, and a high-wavenumber component, $\mathbf{r}$,

$$\mathbf{m} = \mathbf{b} + \mathbf{r}. \tag{9}$$

We refer to $\mathbf{b}$ and $\mathbf{r}$ as the *background* and *reflectivity* components, respectively. Now we perturb each component,

$$\mathbf{m} = \mathbf{b}_0 + \mathbf{\Delta b} + \mathbf{r}_0 + \mathbf{\Delta r}, \tag{10}$$

where $\mathbf{b}_0$ constitutes the *most background* model, $\mathbf{\Delta b}$ is the *perturbation in the background* model, $\mathbf{r}_0$ is the *background reflectivity*, and $\mathbf{\Delta r}$ is the *perturbation in the reflectivity.* In practice, we assume that the model parameters (*e.g.*, slowness) have transitions smooth enough to allowing us to neglect the background reflectivity, hence $\mathbf{r}_0 = \mathbf{0}$. Therefore, the reflectivity image is defined exclusively by $\mathbf{\Delta r}$. For such reason we sometimes refer to it as "the image".

With these considerations we can recast the Newton's equation in such a way that we keep $\mathbf{b}_0$ unchanged during the inversion, and only update the perturbations by setting $\delta\mathbf{m} = \mathbf{\Delta b} + \mathbf{\Delta r}$. Thus, Equation 8 becomes

$$\mathbf{H}(\mathbf{b}_0)[\mathbf{\Delta b} + \mathbf{\Delta r}] = -\nabla\Phi_{FWI}(\mathbf{b}_0). \tag{11}$$

We first interpret the meaning of the right-hand side of Equation 11. Back to Equation 5, let us set $\mathbf{m} = \mathbf{b}_0$,

$$\Phi_{FWI}(\mathbf{b}_0) = \frac{1}{2}\|\mathcal{L}(\mathbf{b}_0) - \mathbf{d}_r\|_2^2, \tag{12}$$

with the corresponding gradient becoming

$$\nabla\Phi_{FWI}(\mathbf{b}_0) = \left[\left.\frac{\partial\mathcal{L}(\mathbf{m})}{\partial\mathbf{m}}\right|_{\mathbf{m}=\mathbf{b}_0}\right]' \left[\mathcal{L}(\mathbf{b}_0) - \mathbf{d}_r\right], \tag{13}$$

where $'$ represents the adjoint. The synthetic data evaluated at the most background model contains transmitted and direct waves, *but no reflections*, while the recorded data contains all the events. Therefore, this data difference corresponds to the negative of the recorded data after removing the transmitted and direct waves, *i.e.*, the negative of the reflection data, $\mathbf{d}_0 = \mathbf{d}_r - \mathcal{L}(\mathbf{b}_0)$. The derivative of the synthetic data evaluated at the most background model is the *Born modeling approximation.* The adjoint of this operator is *reverse-time migration* (RTM) (Barnier and Almomin, 2014). Therefore, the gradient in Equation 13 represents the negative of the reverse-time migration image,

$$\nabla\Phi(\mathbf{b}_0) = -\mathbf{\Delta r}_{mig}, \tag{14}$$

where we represent the migrated image as $\mathbf{\Delta r}_{mig}$ to be consistent with the notation employed throughout this report. The migrated image (and hence the gradient) remains unchanged during the optimization process.

Now we interpret the left-hand side of Equation 11. According to Biondi et al. (2015) we can split the full Hessian $\mathbf{H}$ into the so-called Gauss-Newton Hessian, $\mathbf{H}_{GN}$, and the wave-equation migration velocity analysis (WEMVA) Hessian, $\mathbf{H}_W$:

$$\mathbf{H} = \mathbf{H}_{GN} + \mathbf{H}_W. \tag{15}$$

Substituting Equation 14 and Equation 15 into Equation 11 we obtain

$$(\mathbf{H}_{GN} + \mathbf{H}_W)(\mathbf{\Delta b} + \mathbf{\Delta r}) = \mathbf{\Delta r}_{mig}. \tag{16}$$

We can explicitely apply each Hessian to the perturbations, regrouping terms in a convenient manner:

$$(\mathbf{H}_{GN}\mathbf{\Delta r} + \mathbf{H}_W\mathbf{\Delta b}) + (\mathbf{H}_{GN}\mathbf{\Delta b} + \mathbf{H}_W\mathbf{\Delta r}) = \mathbf{\Delta r}_{mig}. \tag{17}$$

We first analyze the first part of the left-hand side of Equation 17. The first term, $\mathbf{H}_{GN}\mathbf{\Delta r}$, is the Gauss-Newton Hessian applied to the perturbation of the reflectivity. The Gauss-Newton Hessian constitutes the adjoint of the Born modeling operator followed by the corresponding forward operator: $\mathbf{H}_{GN} = \mathbf{L}^T\mathbf{L}$. Fitting this term alone to the migrated image represents conventional LWI in *model space* (Valenciano et al., 2009), to be distinguished from the LWI formulation in *data space* (Nemeth and Schuster, 1999) (both methods are sketched out in Equations 1 to 3). Therefore, this term yields a reflectivity estimation that is more accurate than the conventionally migrated image, in terms of amplitude and seismic resolution. The second term, $\mathbf{H}_W\mathbf{\Delta b}$, is the WEMVA operator applied to the perturbation of the background model, which yields a perturbation of the reflectivity image. Such term is what we need to implement the proposed method.

Now we analyze the second part of the left-hand side of Equation 17. The first term, $\mathbf{H}_{GN}\mathbf{\Delta b}$, is the Gauss-Newton Hessian applied to the perturbation of the background model. Insofar as our objective is to account for reflectivity estimations, we can neglect this term because improving the background model is not part of this objective. The second term, $\mathbf{H}_W\mathbf{\Delta r}$, is the WEMVA operator applied to the perturbation of the reflectivity. This term accounts for second-order scattering effects, such as multiples. We do not consider such effects for LWIVU, therefore we can neglect this term too.

After these considerations, we simplify Equation 17 to obtain

$$\mathbf{H}_{GN}\mathbf{\Delta r} + \mathbf{H}_W\mathbf{\Delta b} = \begin{bmatrix}\mathbf{H}_{GN} & \mathbf{H}_W\end{bmatrix}\begin{bmatrix}\mathbf{\Delta r}\\ \mathbf{\Delta b}\end{bmatrix} \approx \mathbf{\Delta r}_{mig}. \tag{18}$$

From Equation 18 we can cast an optimization problem with the following misfit function,

$$\Phi_1(\mathbf{\Delta r}, \mathbf{\Delta b}; \mathbf{b}_0) = \frac{1}{2}\|\mathbf{H}_{GN}\mathbf{\Delta r} + \mathbf{H}_W\mathbf{\Delta b} - \mathbf{\Delta r}_{mig}\|_2^2, \tag{19}$$

where the optimization parameters are $\mathbf{\Delta r}$ and $\mathbf{\Delta b}$, for a seismic experiment realization with the most background model $\mathbf{b}_0$. Both Hessians are independent of the optimization parameters; they only depend on $\mathbf{b}_0$. Hence, the optimization problem portrayed by Equation 19 is *linear*. We have labeled this misfit function as Φ_1 because an additional constraint is required to update $\mathbf{\Delta b}$ in synergy with the original purpose of producing more focused images. One way to ensure this happens is enforcing the maximization of stacking power, or conversely, the minimization of the *negative* of stacking power. On these grounds we include an additional term in the misfit function,

$$\Phi_2(\mathbf{\Delta b}; \mathbf{b}_0) = -\frac{\lambda^2}{2}\|\mathbf{H}_W\mathbf{\Delta b} + \mathbf{\Delta r}_{mig}\|_2^2, \tag{20}$$

where λ is a parameter that allows us to control the level of stacking power maximization. In order to analyze the interaction between the misfit functions Φ_1 and Φ_2, we express Equation 19 as an optimization problem aimed at fitting the reflectivity model to an improved migrated image $\widehat{\mathbf{\Delta r}}_{mig}(\mathbf{\Delta b})$, such that

$$\widehat{\mathbf{\Delta r}}_{mig}(\mathbf{\Delta b}) = \mathbf{\Delta r}_{mig} - \mathbf{H}_W \mathbf{\Delta b}, \tag{21}$$

therefore Equation 19 becomes

$$\Phi_1(\mathbf{\Delta r}, \mathbf{\Delta b}; \mathbf{b}_0) = \frac{1}{2} \| \mathbf{H}_{GN} \mathbf{\Delta r} - \widehat{\mathbf{\Delta r}}_{mig}(\mathbf{\Delta b}) \|_2^2, \tag{22}$$

According to this definition, once $\mathbf{\Delta b}$ satisfies the misfit function, Φ_2 (thus maximizing the stacking power), it should yield an improved migrated image, $\widehat{\mathbf{\Delta r}}_{mig}$. However, the maximizing effect gained by $\mathbf{\Delta b}$ precisely *removes* such correction from such migrated image, because of the *minus* sign in Equation 21. Such a scheme would harm the reflectivity estimation. Hence, it becomes logical to re-define the improved migrated image as the term to be maximized inside the norm in Equation 20,

$$\widehat{\mathbf{\Delta r}}_{mig}(\mathbf{\Delta b}) = \mathbf{\Delta r}_{mig} + \mathbf{H}_W \mathbf{\Delta b}, \tag{23}$$

thus the LWIVU misfit function becomes

$$\begin{aligned} \Phi(\mathbf{\Delta r}, \mathbf{\Delta b}; \mathbf{b}_0) &= \Phi_1(\mathbf{\Delta r}, \mathbf{\Delta b}; \mathbf{b}_0) + \Phi_2(\mathbf{\Delta b}; \mathbf{b}_0) \\ &= \frac{1}{2} \| \mathbf{H}_{GN} \mathbf{\Delta r} - \widehat{\mathbf{\Delta r}}_{mig}(\mathbf{\Delta b}) \|_2^2 - \frac{\lambda^2}{2} \| \widehat{\mathbf{\Delta r}}_{mig}(\mathbf{\Delta b}) \|_2^2. \end{aligned} \tag{24}$$

Let us explicitely express Equation 24 in terms of $\mathbf{H\Delta b}$:

$$\Phi(\mathbf{\Delta r}, \mathbf{\Delta b}; \mathbf{b}_0) = \frac{1}{2} \| \mathbf{H}_{GN} \mathbf{\Delta r} - (\mathbf{H}_W \mathbf{\Delta b} + \mathbf{\Delta r}_{mig}) \|_2^2 - \frac{\lambda^2}{2} \| \mathbf{H}_W \mathbf{\Delta b} + \mathbf{\Delta r}_{mig} \|_2^2. \tag{25}$$

In order to implement gradient-based optimization schemes we obtain the gradient of the misfit function in Equation 25. Deriving with respect to the model parameters $\mathbf{\Delta r}$ and $\mathbf{\Delta b}$,

$$\nabla_{\mathbf{\Delta r}} \Phi = \mathbf{H}'_{GN} (\mathbf{H}_{GN} \mathbf{\Delta r} - \mathbf{H}_W \mathbf{\Delta b} - \mathbf{\Delta r}_{mig}), \tag{26}$$

$$\nabla_{\mathbf{\Delta b}} \Phi = \mathbf{H}'_W (\mathbf{H}_{GN} \mathbf{\Delta r} - \mathbf{H}_W \mathbf{\Delta b} - \mathbf{\Delta r}_{mig}) - \lambda \mathbf{H}'_W \Big[\lambda (\mathbf{H}_W \mathbf{\Delta b} + \mathbf{\Delta r}_{mig}) \Big], \tag{27}$$

the total gradient becomes

$$\nabla \Phi = \begin{bmatrix} \nabla_{\mathbf{\Delta r}} \Phi \\ \nabla_{\mathbf{\Delta b}} \Phi \end{bmatrix} = \begin{bmatrix} \mathbf{H}'_{GN} & \mathbf{0} \\ -\mathbf{H}'_W & -\lambda \mathbf{H}'_W \end{bmatrix} \begin{bmatrix} \mathbf{H}_{GN} \mathbf{\Delta r} - \mathbf{H}_W \mathbf{\Delta b} - \mathbf{\Delta r}_{mig} \\ \lambda (\mathbf{H}_W \mathbf{\Delta b} + \mathbf{\Delta r}_{mig}). \end{bmatrix} \tag{28}$$

The *adjoint* of LWIVU is the matrix-like operator at the right-hand side of Equation 28,

$$\begin{bmatrix} \mathbf{H}'_{GN} & \mathbf{0} \\ -\mathbf{H}'_W & -\lambda \mathbf{H}'_W \end{bmatrix}, \tag{29}$$

which is applied to the residuals vector. Note that such residuals correspond to the *fitting goals*, i.e., the optimization terms in Equation 25. The adjoint of the LWIVU adjoint operator constitutes the LWIVU *forward* operator:

$$\begin{bmatrix} \mathbf{H}_{GN} & -\mathbf{H}_W \\ \mathbf{0} & -\lambda \mathbf{H}_W \end{bmatrix}. \tag{30}$$

With the forward and adjoint operators we can set forth LWIVU as a gradient-based optimization scheme. We implement the optimization using the *conjugate directions method* (Claerbout, 2014). We discuss such implementation in the Appendix A. In the Appendix B we derive the WEMVA optimization problem using maximization of the stacking power, in an effort to provide the reader with additional tools for understanding the LWIVU derivation.

SYNTHETIC EXAMPLES: LWI VS. LWIVU

For the sake of illustration we implement LWIVU in a 2-D synthetic dataset obtained from a two-layer subsurface model with a Gaussian anomaly (Figure 1). The model space consists of 151×461 grid-points in depth (z) and distance (x), respectively, plus a 48-point haloing zone for the tapering boundary conditions. Grid spacing is $\Delta x = \Delta z = 20$ m. The slowness square of the upper and lower layers are 2.5×10^{-7} s^2/m^2 (2000 m/s) and 1.6×10^{-7} s^2/m^2 (2500 m/s), respectively. The Gaussian anomaly (Figure 2) has a minimum value of -2.3×10^{-8} s^2/m^2. It is added to the first layer. For this experimental setup, the initial slowness squared model consists of the two-layer model *without* the anomaly (Figure 3). This model corresponds to the most background model, $\mathbf{b}_0$, whereas the anomaly constitutes the perturbation of the background model, $\boldsymbol{\Delta}\mathbf{b}$.

We obtained synthetic data by non-linear modeling using the true model and a constant model (the latter consists of the slowness of the upper layer only), and subtracted them to remove direct waves and only preserve reflections. The acquisition geometry consists of 31 sources spaced every 300 m with the first source at $x = 0$. Every source corresponds to 481 receivers spaced every 20 m. Source and receivers lie at the surface, $z = 0$.

We obtained the migrated image, $\boldsymbol{\Delta}\mathbf{r}_{mig}$, by applying RTM to the data with the initial model (Figure 2(d)). The Gauss-Newton Hessian and WEMVA Hessian are calculated with the most background model, $\mathbf{b}_0$. For the experiments shown in this report we only use the diagonal of the Gauss-Newton Hessian matrix, yielding a correction in the amplitude but not a resolution increase.

In the following plots we compare the LWI reflectivity (left panels) with the LWIVU reflectivity (right panels), and the true slowness squared perturbation (left panels) with the LWIVU perturbation (right panels), for different λ values. We obtained the LWI result in model space according to Equation 4, which becomes $\Phi(\boldsymbol{\Delta}\mathbf{r}) = \frac{1}{2}\|\mathbf{H}_{GN}\boldsymbol{\Delta}\mathbf{r} - \boldsymbol{\Delta}\mathbf{r}_{mig}\|_2^2$ with the notation used in the theoretical section. For better appreciation we zoomed in on the reflectivity panels. We applied no clippling to the reflectivity panels. In the slowness squared perturbations panels we clipped to the minimum of the true anomaly. All the examples were run for 10 iterations, unless otherwise specified.

Figure 5 shows the reflectivity comparison for $\lambda = 0$, *i.e.*, without stacking power maximization. Note the amplitude gap at the center of the reflector in both panels caused by the incorrect slowness model. The excess in slowness square when ignoring the anomaly is too small to produce kinematic errors (*e.g.*, pull-ups), thus only amplitudes are affected. Two linear events cross at the center of the reflector gap. They are the smearing effect of conventional migration, and off-diagonal elements of the Gauss-Newton Hessian are needed to collapse such artifacts. LWIVU with $\lambda = 0$ did not recover the amplitude at the gap. Moreover, the perturbation in slowness squared is updated with the wrong polarity (positive)

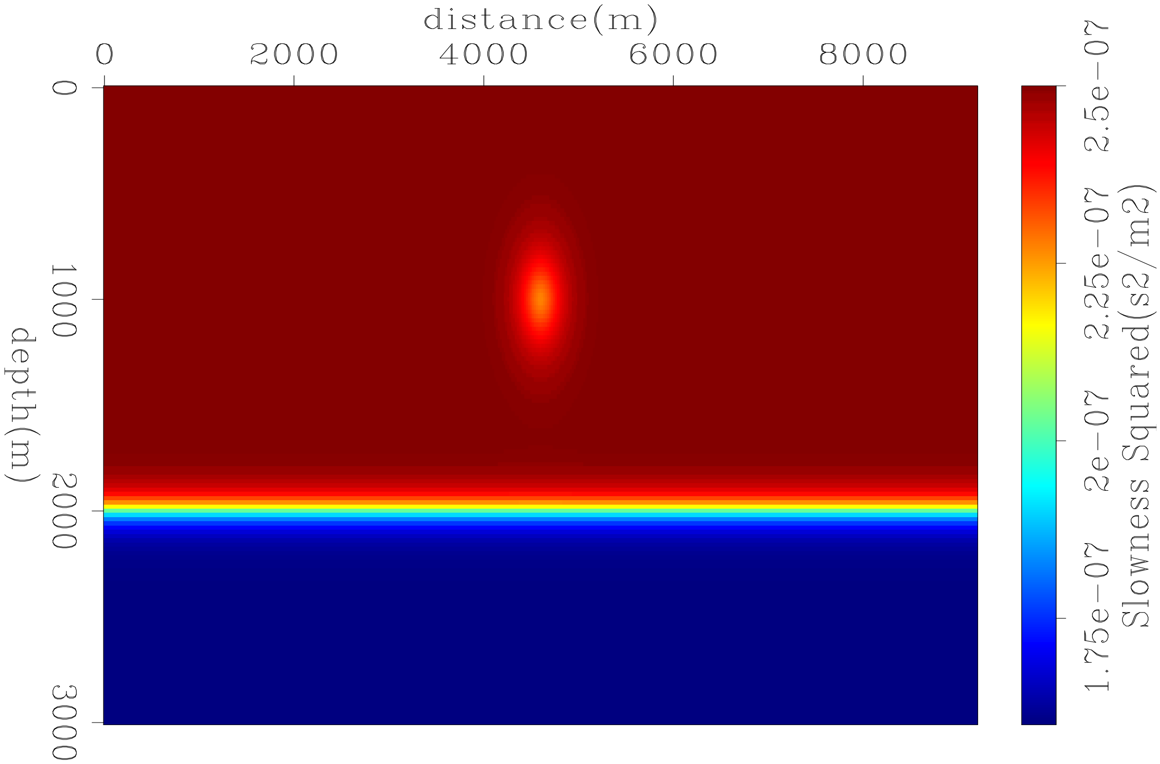


Figure 1: Slowness squared model with Gaussian anomaly. This model constitutes the real model [**ER**]. alejandro1/. mod-gauss

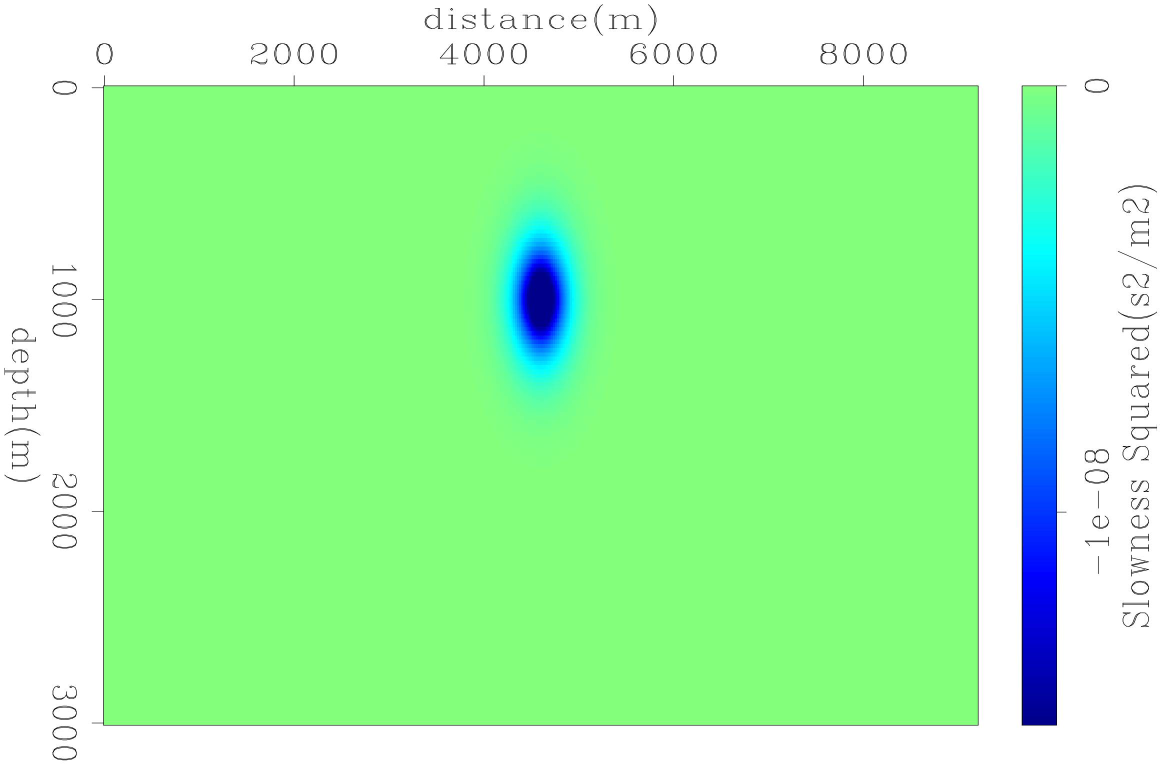


Figure 2: Slowness squared perturbation, $\mathbf{\Delta b}$: Gaussian anomaly [**ER**]. alejandro1/. gauss

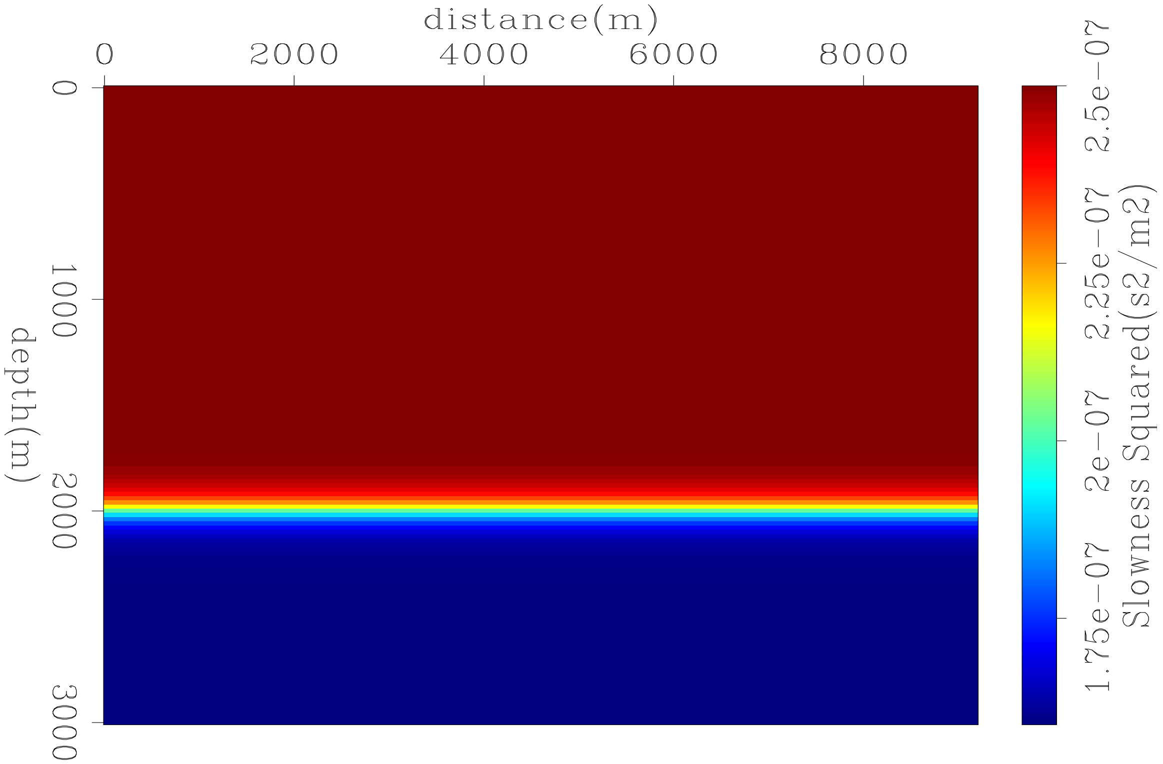


Figure 3: Slowness squared model without Gaussian anomaly. This model constitutes the most background model, $\mathbf{b}_0$ [**ER**]. alejandro1/. nogauss

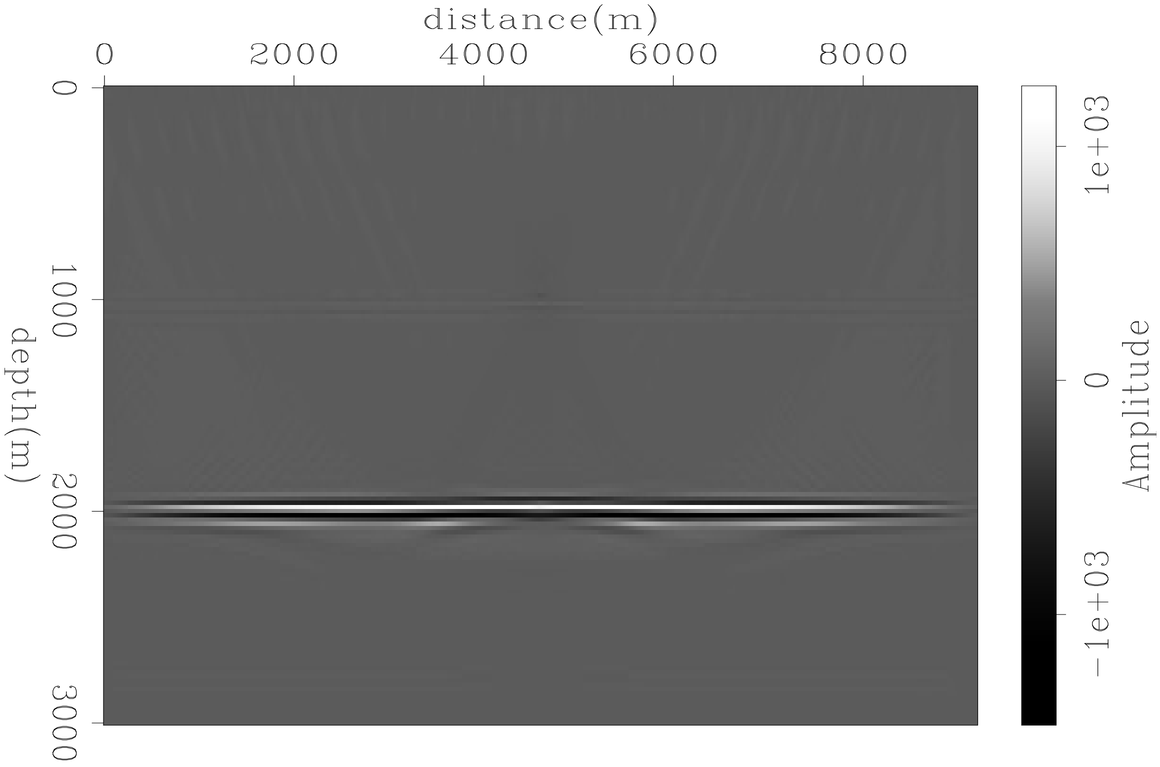


Figure 4: RTM image, $\mathbf{\Delta r}_{mig}$ [**CR**]. alejandro1/. rtm

(Figure 6). The corresponding "improved" migrated image, $\widehat{\mathbf{\Delta r}}_{mig}$, has the gap reinforced, and the reflectivity is fitted to such undesired image.

Figure 7 shows the reflectivity comparison for $\lambda = 0.5$. The results are very similar to the previous case. However, the perturbation in slowness squared (Figure 8) is less positive than that of $\lambda = 0$ (Figure 6).

Figure 9 shows the reflectivity comparison for $\lambda = 1.0$. There is still no apparent change with respect to the previous cases, but the slowness squared perturbation flipped the sign and became negative (Figure 10). Imposing stacking power maximization pushes the perturbation of the image to fill the gap in the migrated image, which demands a negative $\mathbf{\Delta b}$. The reflectivity is fitted to this improved migrated image. Yet, 10 iterations are not enough to match the true slowness squared with $\lambda = 1.0$, then the improved migrated image does not have the gap filled, and as a consequence, neither does the reflectivity. Thus, we extended the inversion to 16 iterations. Note in Figure 11 that the gap previously observed in the right panel of Figure 9 has been healed. The shaded zone still observed at the reflector center can be the crossing linear events. The background in slowness squared in Figure 12 now satisfactorily resembles the true perturbation.

Figure 13 shows the reflectivity comparison for $\lambda = 1.5$. The time the gap has been healed, and the overall amplitude is increased. However, the perturbation appears to be slightly surpassing the true anomaly (Figure 14). Note the increase in amplitude of the subsamplig artifacts in the shallow part and the reflector leaking into the perturbation image. They appear not to be influencing the reflectivity. Using less iterations we can better fit the perturbation in slowness and reduce the artifacts, but then the gap is not completely healed.

Figure 15 shows the reflectivity comparison for $\lambda = 2.5$. The gap is eliminated, but the amplitude appears to be over-corrected (note the bias towards positive values in the scalebar). As expected, $\mathbf{\Delta b}$ surpasses the true slowness (Figure 16) further than $\lambda = 1.5$. Better results are obtained at the eighth iteration (Figure 17, note the unbiased scalebar), although $\mathbf{\Delta b}$ remains overestimated (Figure 18). Despite the over-correction, the overall amplitude decreased compared to $\lambda = 1.5$. The artifacts in $\mathbf{\Delta b}$ increased with respect to $\lambda = 1.5$ in both cases.

Finally, Figure 19 and Figure 20 are the comparisons of reflectivity and perturbation in slowness squared using $\lambda = 5.0$. The gap is over-corrected, whereas the amplitude significantly diminishes further away from the reflector center. In addition, the perturbation in slowness squared has been laterally compressed. The positive surrounding halo, the subsampling artifacts, and the leaked reflector are exacerbated compared to lower λ values. The reflectivity amplitudes can be impacted by such effects. The energy imaging condition (Rocha et al., 2016) can potentially attenuate the leaked reflectivity, whereas the subsampling artifacts can be attenuated by addressing the model null space.

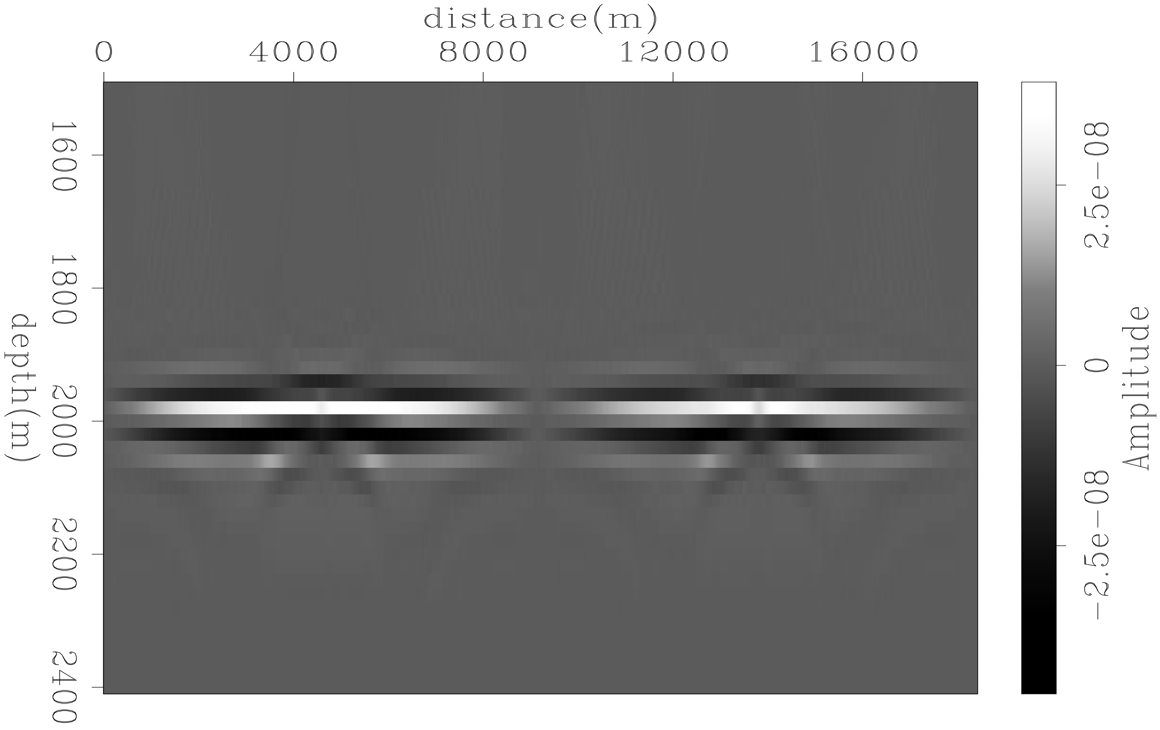


Figure 5: Reflectivity comparison, $\lambda = 0$. *Left*: LWI reflectivity; *Right*: LWIVU reflectivity [**CR**]. alejandro1/. Fig-R-0-0

COMPUTATIONAL COST OF LWIVU

LWIVU demands more wavefield propagations than LWI because of the inclusion of WEMVA into the process. We compare LWI and LWIVU flowcharts to quantify them. The optimization scheme is the conjugate directions method (Appendix A). We assume reversible propagations by using random boundary conditions (Clapp, 2009, 2010), which massively reduce storage demand at the cost of extra-propagations. The arrows accompanying the source/receiver wavefield symbols signify forward propagation (right-pointing arrow) and backward propagation (left-pointing arrow) in time.

Figure 21 shows a flowchart of LWI performed in data space (in model space is trivial; LWI is virtually instantaneous once the Gauss-Newton Hessian has been pre-computed). We perform source wavefield and residual wavefield backward propagations in order to compute the gradient each iteration. In other words, the adjoint of Born modeling or RTM (a forward propagation of the source wavefield is required before the iterative scheme). Subsequent projection of the gradient into data space requires Born modeling. Thus, we perform forward propagation of the source wavefield and the source scattered wavefield, using the gradient as entry for reflectivity. The total number of propagations is $1 + 4n_{iter}$, where n_{iter} is the number of iterations.

Figure 22 shows a flowchart of LWIVU. We precompute the Gauss-Newton Hessian, e.g., using point-spread functions (Fletcher et al., 2016), and apply it on demand. Now the gradient consists of two members: one relates to the reflectivity and the other relates

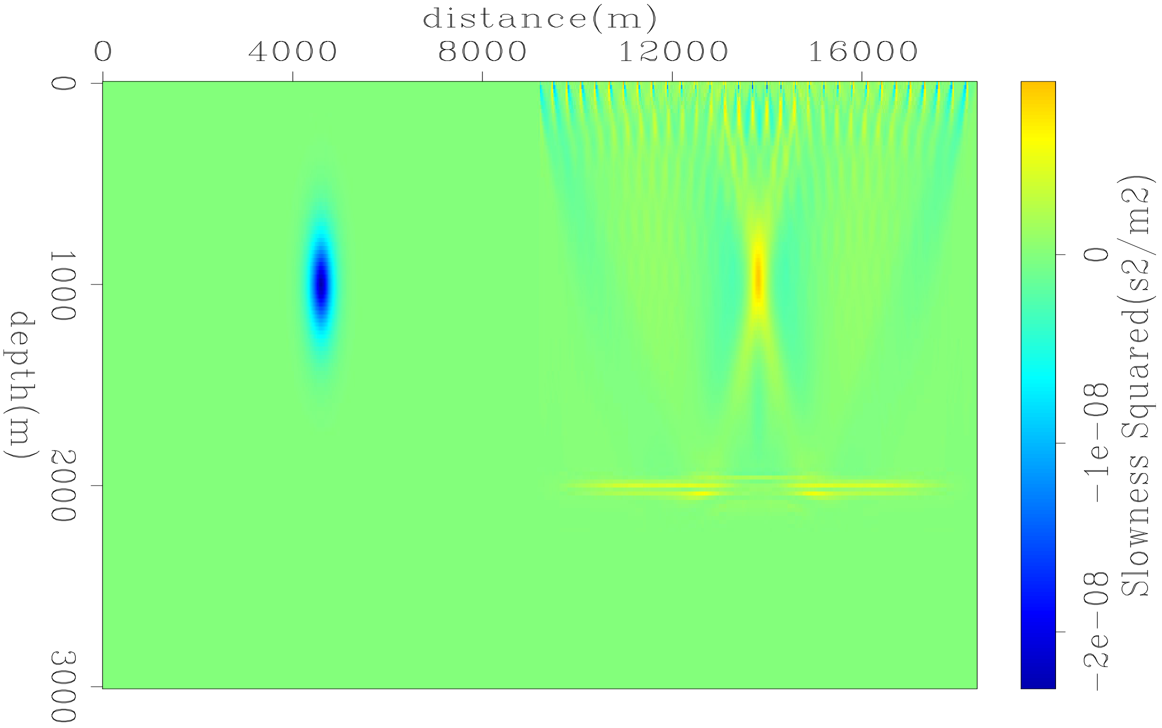


Figure 6: Perturbation in slowness squared, $\lambda = 0$. *Left*: True anomaly; *Right*: LWIVU perturbation [**CR**]. alejandro1/. Fig-B-0-0

to the perturbation of the background model (see Appendix A for details). The former consists of applying the Gauss-Newton Hessian to the appropriate residual, not requiring wavefield propagations. The latter consists of applying the WEMVA Hessian to a linear combination of the residuals. Here WEMVA requires four backward wavefield propagations of the source wavefield and the receiver wavefield, among the souce scattered wavefield and the receiver scattered wavefield, using the residuals as entries for perturbations in the image (constituting the adjoing of WEMVA). For the projection of the gradient into data space we apply WEMVA again. We forward propagate the source wavefield and the receiver wavefield, among the source scattered wavefield and the receiver scattered wavefield using the appropriate member of the gradient as entry for perturbations of the background model (constituting forward WEMVA). Accounting for the initial forward propagation of the source wavefield, the total number of propagations is $1 + 8n_{iter}$, *i.e.* twice as much as in LWIVU, and excluding pre-computation of the Gauss-Newton Hessian.

From this analysis we recognize the computational demands as the main caveat of LWIVU. Nonetheless, we envision a situation where this method can be applicable. In reservoir characterization, it can be advantageous to perform LWI in model space in order to take advantage of amplitude variations with enhanced resolution, for better identification of oil-targets. We typically use small seismic volumes encompassing only the oil field of interest, so expensive iterative-based imaging methods become affordable. In case of small velocity errors impacting the amplitudes, we can implement LWIVU to rectify the reflectivity, without re-computing the Gauss-Newton Hessian.

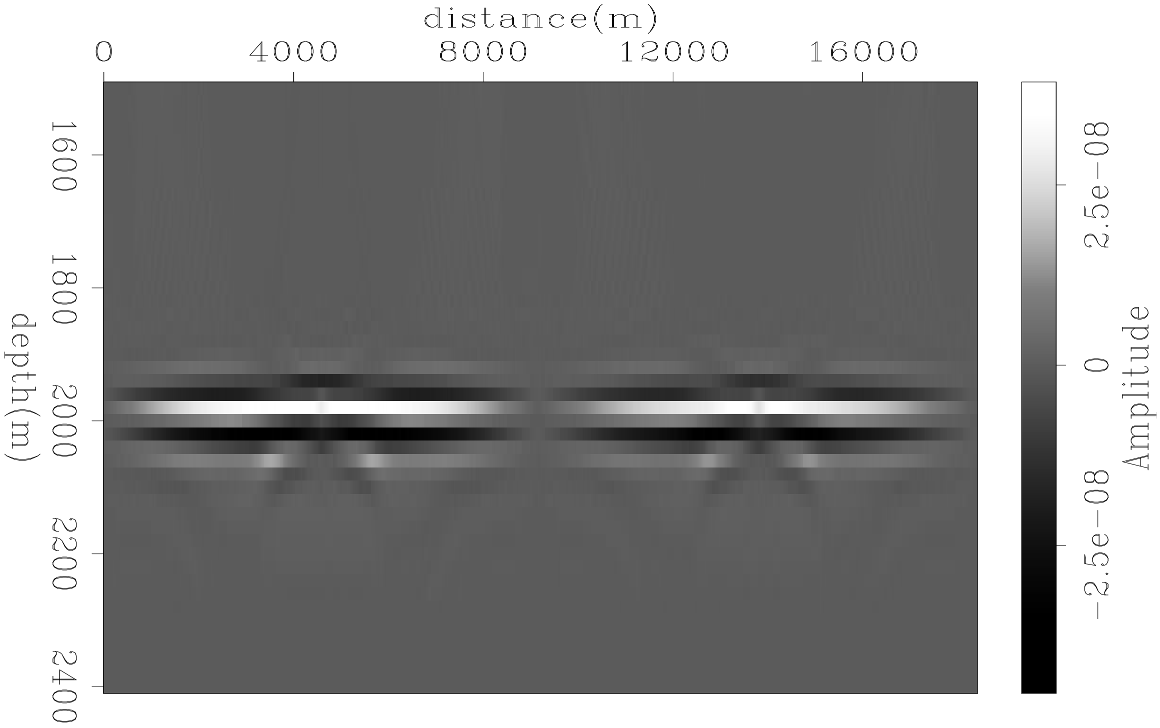


Figure 7: Reflectivity comparison, $\lambda = 0.5$. *Left*: LWI reflectivity; *Right*: LWIVU reflectivity [**CR**]. alejandro1/. Fig-R-0-5

CONCLUSIONS

We revisited the LWIVU theory to make the improved image (after stacking power maximization) the target image for the reflectivity estimation. The synthetic examples demonstrate the potential of this method to correct amplitude inaccuracies derived from the cumulative effect of velocity (or slowness) errors. The main disadvantage is the computational demands, which are about twice compared to LWI, besides requiring the pre-computation of the Gauss-Newton Hessian. However, we foresee the applicability of LWIVU in detailed exploration works, such as reservoir characterization, where accurately addressing amplitude variations is vital for identification and delimitation of oil-bearing rock facies.

ACKNOWLEDGEMENT

We thank the SEP sponsors for their continuous support. Alejandro Cabrales thanks Petróleos Mexicanos for sponsoring his studies.

APPENDIX A: THE CONJUGATE DIRECTIONS METHOD AND LWIVU

The software that implements the gradient-based optimization is based on the conjugate directions method (Claerbout, 2014), a "cousin" of the more popular conjugate gradient

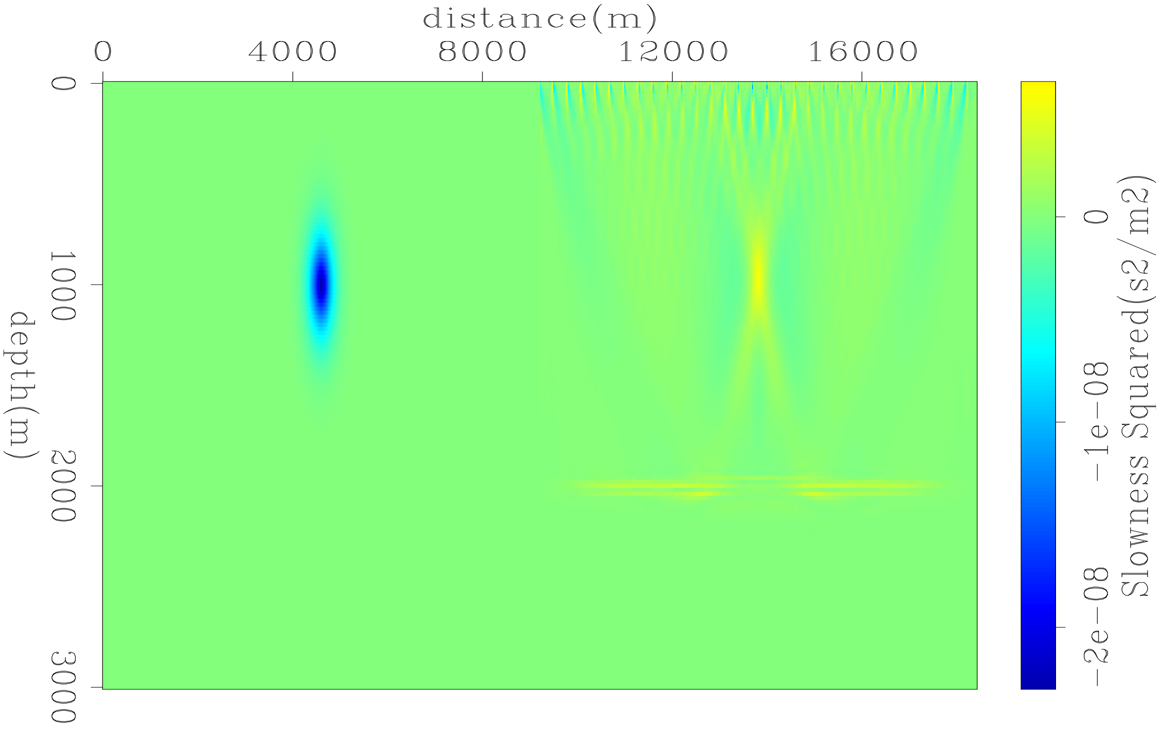


Figure 8: Perturbation in slowness squared, $\lambda = 0.5$. *Left*: True anomaly; *Right*: LWIVU perturbation [**CR**]. alejandro1/. Fig-B-0-5

method (Hestenes and Stiefel, 1952). In this Appendix I illustrate its implementation directly with LWIVU.

I initialize the method with an starting model. In linear problems it is often practical to set this model to zero. The model space consists of the perturbation of reflectivity and perturbation of the background, so the initial model is

$$\begin{bmatrix} \mathbf{\Delta r}_0 \\ \mathbf{\Delta b}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \tag{A-1}$$

Then I compute the first residual, $\mathbf{f}_0$, by using the fitting goals in Equation 25 evaluated at the initial model:

$$\mathbf{f}_0 = \begin{bmatrix} \mathbf{f}_0^{\mathbf{\Delta r}_1} \\ \mathbf{f}_0^{\mathbf{\Delta r}_2} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{GN}\mathbf{\Delta r}_0 - \mathbf{H}_W\mathbf{\Delta b}_0 - \mathbf{\Delta r}_{mig} \\ \lambda(\mathbf{H}_W\mathbf{\Delta b}_0 + \mathbf{\Delta r}_{mig}) \end{bmatrix} = \begin{bmatrix} -\mathbf{\Delta r}_{mig} \\ \lambda\mathbf{\Delta r}_{mig} \end{bmatrix} \tag{A-2}$$

The data space is the conventionally-migrated image space. Thus, the two components of the residual, $\mathbf{f}_0^{\mathbf{\Delta r}_1}$ and $\mathbf{f}_0^{\mathbf{\Delta r}_2}$, constitute migrated-like images.

Now I can set forth the iterative process, calculating the gradient at iteration k using the adjoind operator (Equation 28),

$$\mathbf{g}_k = \begin{bmatrix} \mathbf{g}_k^{\mathbf{\Delta r}} \\ \mathbf{g}_k^{\mathbf{\Delta b}} \end{bmatrix} = \begin{bmatrix} \mathbf{H}'_{GN} & \mathbf{0} \\ -\mathbf{H}'_W & -\lambda\mathbf{H}'_W \end{bmatrix} \begin{bmatrix} \mathbf{f}_k^{\mathbf{\Delta r}_1} \\ \mathbf{f}_k^{\mathbf{\Delta r}_2} \end{bmatrix} \tag{A-3}$$

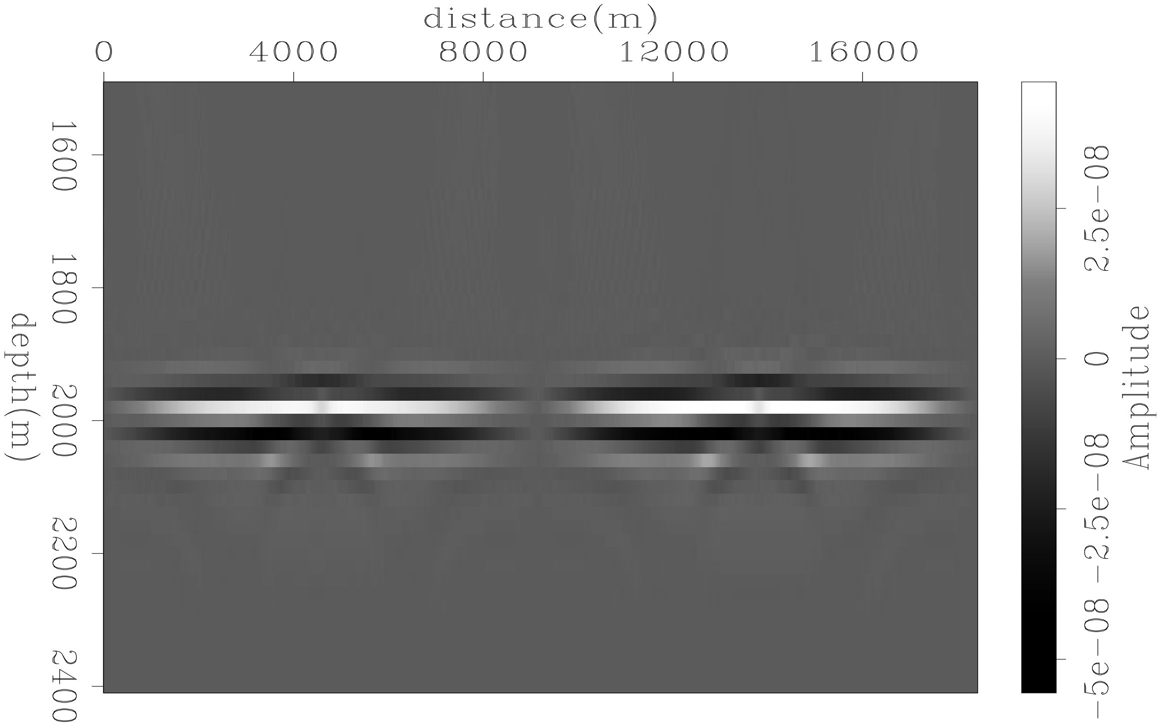


Figure 9: Reflectivity comparison, $\lambda = 1.0$. *Left*: LWI reflectivity; *Right*: LWIVU reflectivity [**CR**]. alejandro1/. Fig-R-1-0

and then projecting such gradient onto the data space using the forward operator (Equation 30),

$$\mathbf{G}_k = \begin{bmatrix} \mathbf{G}_k^{\boldsymbol{\Delta}\mathbf{r}_1} \\ \mathbf{G}_k^{\boldsymbol{\Delta}\mathbf{r}_2} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{GN} & -\mathbf{H}_W \\ \mathbf{0} & -\lambda \mathbf{H}_W \end{bmatrix} \begin{bmatrix} \mathbf{g}_k^{\boldsymbol{\Delta}\mathbf{r}} \\ \mathbf{g}_k^{\boldsymbol{\Delta}\mathbf{b}} \end{bmatrix}. \tag{A-4}$$

What follows next is determining the model update and the residual update. They are given respectively by

$$\begin{bmatrix} \delta \boldsymbol{\Delta}\mathbf{r}_{k+1} \\ \delta \boldsymbol{\Delta}\mathbf{b}_{k+1} \end{bmatrix} = \alpha \begin{bmatrix} \mathbf{g}_k^{\boldsymbol{\Delta}\mathbf{r}} \\ \mathbf{g}_k^{\boldsymbol{\Delta}\mathbf{b}} \end{bmatrix} + \beta \begin{bmatrix} \delta \boldsymbol{\Delta}\mathbf{r}_k \\ \delta \boldsymbol{\Delta}\mathbf{b}_k \end{bmatrix} \tag{A-5}$$

and

$$\begin{bmatrix} \delta \mathbf{f}_{k+1}^{\boldsymbol{\Delta}\mathbf{r}_1} \\ \delta \mathbf{f}_{k+1}^{\boldsymbol{\Delta}\mathbf{r}_2} \end{bmatrix} = \alpha \begin{bmatrix} \mathbf{G}_k^{\boldsymbol{\Delta}\mathbf{r}_1} \\ \mathbf{G}_k^{\boldsymbol{\Delta}\mathbf{r}_2} \end{bmatrix} + \beta \begin{bmatrix} \delta \mathbf{f}_k^{\boldsymbol{\Delta}\mathbf{r}_1} \\ \delta \mathbf{f}_k^{\boldsymbol{\Delta}\mathbf{r}_2} \end{bmatrix}, \tag{A-6}$$

where α and β are parameters that define the search plane that minimizes the new residual (Claerbout, 2014). Finally, I update the model and the residual:

$$\begin{bmatrix} \boldsymbol{\Delta}\mathbf{r}_{k+1} \\ \boldsymbol{\Delta}\mathbf{b}_{k+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Delta}\mathbf{r}_k \\ \boldsymbol{\Delta}\mathbf{b}_k \end{bmatrix} + \begin{bmatrix} \delta \boldsymbol{\Delta}\mathbf{r}_{k+1} \\ \delta \boldsymbol{\Delta}\mathbf{b}_{k+1} \end{bmatrix}, \tag{A-7}$$

$$\begin{bmatrix} \mathbf{f}_{k+1}^{\boldsymbol{\Delta}\mathbf{r}_1} \\ \mathbf{f}_{k+1}^{\boldsymbol{\Delta}\mathbf{r}_2} \end{bmatrix} = \begin{bmatrix} \mathbf{f}_k^{\boldsymbol{\Delta}\mathbf{r}_1} \\ \mathbf{f}_k^{\boldsymbol{\Delta}\mathbf{r}_2} \end{bmatrix} + \begin{bmatrix} \delta \mathbf{f}_{k+1}^{\boldsymbol{\Delta}\mathbf{r}_1} \\ \delta \mathbf{f}_{k+1}^{\boldsymbol{\Delta}\mathbf{r}_2} \end{bmatrix}, \tag{A-8}$$

and proceed with the next iteration.

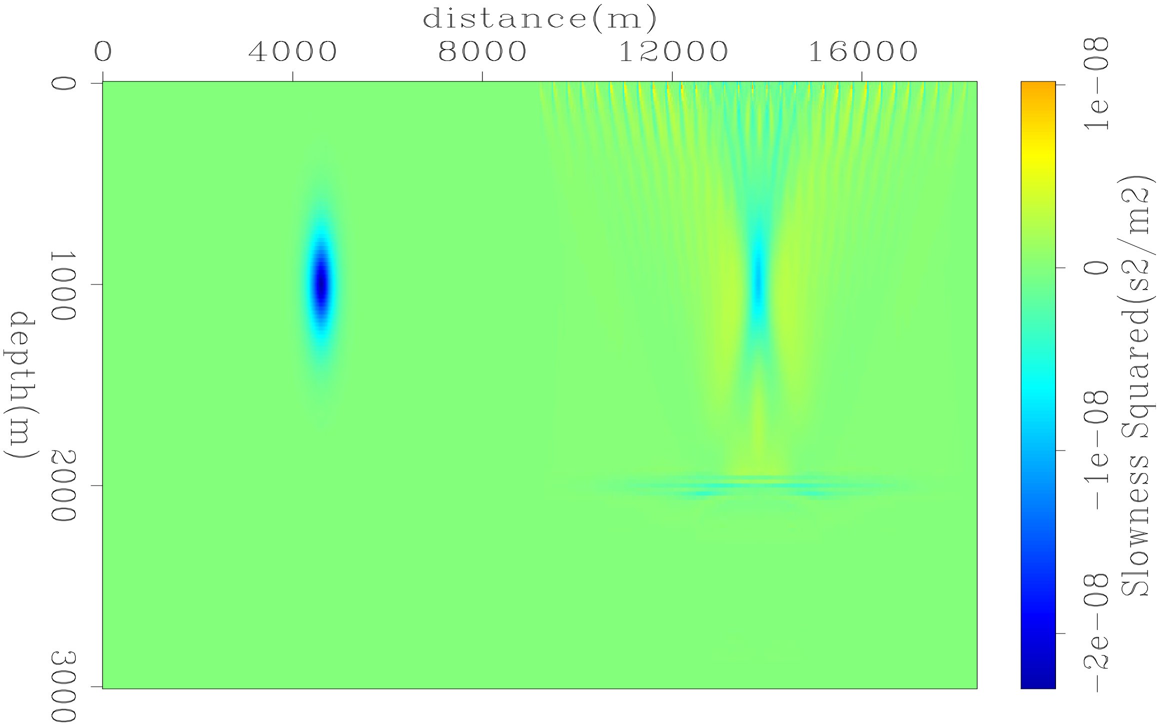


Figure 10: Perturbation in slowness squared, $\lambda = 1.0$. *Left*: True anomaly; *Right*: LWIVU perturbation [**CR**]. alejandro1/. Fig-B-1-0

APPENDIX B: WEMVA INVERSION USING STACKING POWER MAXIMIZATION

In this Appendix I derive the WEMVA inversion process using the maximization of the stacking power as a focusing operator. This derivation is important to have a better understanding of how LWIVU was derived, for maximizing a function via minimizing its negation is tricky. Thus, I explain with special thoroughness through this section.

The general expression for the WEMVA misfit function is (Biondi, 2006)

$$\Phi(\mathbf{s}^2) = \frac{1}{2}\|\mathbf{\Delta r}_{mig}(\mathbf{s}^2) - \mathbf{F}\left[\mathbf{\Delta r}_{mig}(\mathbf{s}^2)\right]\|_2^2, \tag{B-1}$$

where $\mathbf{s}^2$ is the slowness squared field or background model, and $\mathbf{F}$ is a focusing operator applied to the migrated image. For the sake of consistency I keep the notation for the migrated image, $\mathbf{\Delta r}_{mig}$, the same as in the theoretical section.

The focusing operator enforces the correction of the migrated image *without changing the background model*. Such corrected image will become the *target image* that the inversion process fits, now updating the background model. There are several focusing operators. The most popular are perhaps the *differential semblance optimization* (DSO) (Symes and Carazzone, 1991) and the maximization of the stacking power. The former enforces focusing by flattening the angle-domain common-image gathers (ADCIG), whereas the latter seeks for maximum power of the stack section, presumably occurring when offset-domain common-

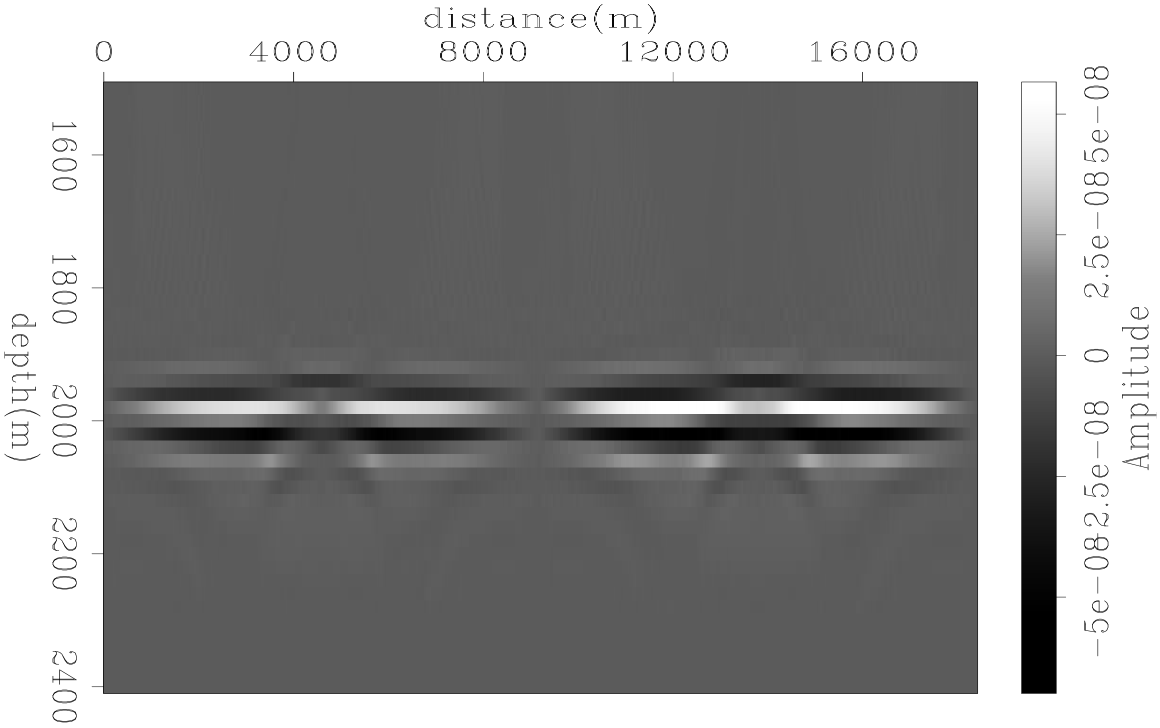


Figure 11: Reflectivity comparison, $\lambda = 1.0$. *Left*: LWI reflectivity; *Right*: LWIVU reflectivity. The number of iterations was increased to 16 [**CR**]. alejandro1/. Fig-R-1-0a

image gathers (ODCIG) focus, an it is equivalent to flattening ADCIG. Maximazing the stacking power is known to suffer from cycle skipping when velocity errors are big, but in LWIVU we assume that velocity inaccuracies are rather small, therefore it becomes applicable.

Maximization of the stacking power is achieved using $\mathbf{F} = \mathbf{I} + \mathbf{S}$ as a focusing operator, where $\mathbf{I}$ is the identity operator and $\mathbf{S}$ is the stacking operator. Substituting into Equation B-1, we obtain

$$\Phi(\mathbf{s}^2) = \frac{1}{2}\|\mathbf{\Delta r}_{mig}(\mathbf{s}^2) - \left[\mathbf{\Delta r}_{mig}(\mathbf{s}^2) + \mathbf{S\Delta r}_{mig}(\mathbf{s}^2)\right]\|_2^2 = \frac{1}{2}\| - \mathbf{S\Delta r}_{mig}(\mathbf{s}^2)]\|_2^2 \qquad \text{(B-2)}$$

The minus sign is lost if we simply perform the algebra, yielding an ordinary minimization. In order to achieve the maximization we move the negative sign out of the absolute value (e.g. Tang, 2011a),

$$\Phi(\mathbf{s}^2) = \frac{1}{2}\|\mathbf{\Delta r}_{mig}(\mathbf{s}^2) - \left[\mathbf{\Delta r}_{mig}(\mathbf{s}^2) + \mathbf{S\Delta r}_{mig}(\mathbf{s}^2)\right]\|_2^2 = -\frac{1}{2}\|\mathbf{S\Delta r}_{mig}(\mathbf{s}^2)]\|_2^2 \qquad \text{(B-3)}$$

In the case of zero subsurface offset, the stacking operator becomes the identity operator, thus

$$\Phi(\mathbf{s}^2) = -\frac{1}{2}\|\mathbf{\Delta r}_{mig}(\mathbf{s}^2)]\|_2^2 \qquad \text{(B-4)}$$

We can linearize this problem by assuming that $\mathbf{s}^2 = \mathbf{s}_0^2 + \mathbf{\Delta s}^2$, where $\mathbf{s}_0^2$ is the initial slowness squared model (analogous to the most background model), and $\mathbf{\Delta s}^2$ is a perturbation in

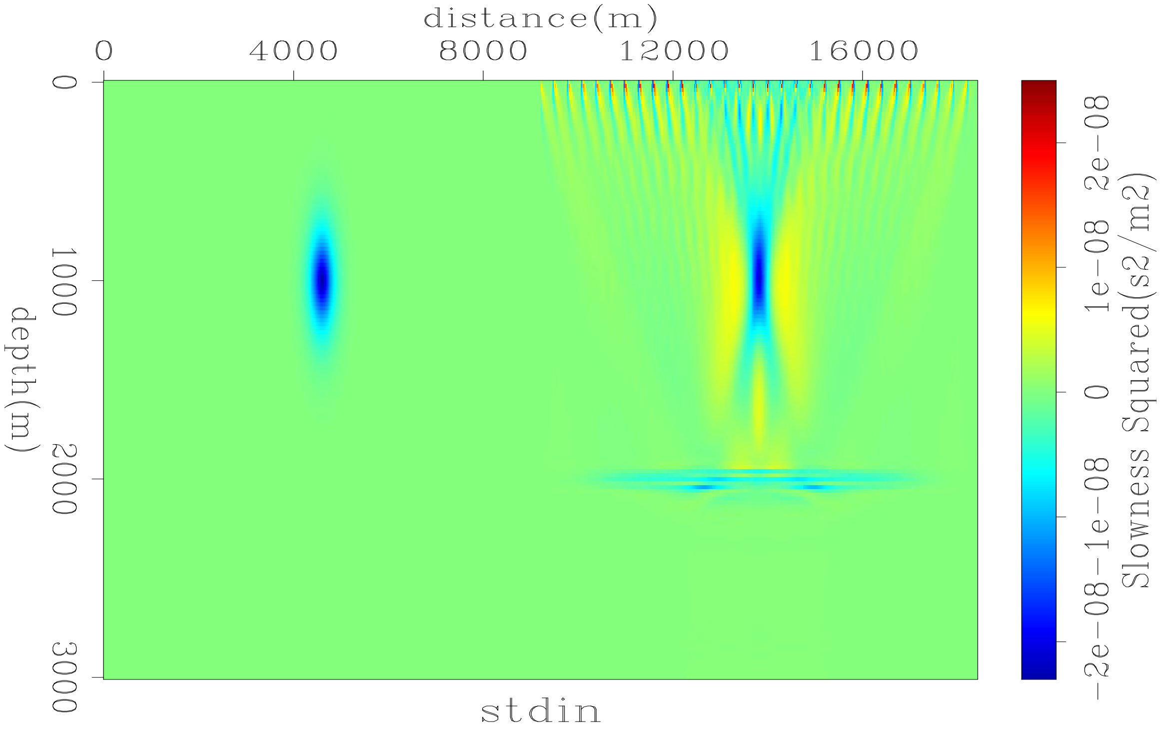


Figure 12: Perturbation in slowness squared, $\lambda = 1.0$. *Left*: True anomaly; *Right*: LWIVU perturbation. The number of iterations was increased to 16 [**CR**]. alejandro1/. Fig-B-1-0a

such initial model. Thus, I can expand the migrated image around $\mathbf{s}_0^2$, with $\mathbf{s}^2 = \mathbf{s}_0^2 + \mathbf{\Delta s}^2$ for a small perturbation $\mathbf{\Delta s}^2$,

$$\mathbf{\Delta r}_{mig}(\mathbf{s}^2) \approx \mathbf{\Delta r}_{mig}(\mathbf{s}_0^2) + \left[\frac{\partial \mathbf{\Delta r}_{mig}(\mathbf{s}^2)}{\partial \mathbf{s}^2}|_{\mathbf{s}^2=\mathbf{s}_0^2}\right]\mathbf{\Delta s}^2 = \mathbf{\Delta r}_{mig}(\mathbf{s}_0^2) + \mathbf{W\Delta s}^2, \tag{B-5}$$

where $\mathbf{\Delta r}_{mig}(\mathbf{s}_0^2)$ is the image migrated with $\mathbf{s}_0^2$, and The derivative in Equation B-7 constitutes the WEMVA operator, $\mathbf{W}$ (equivalent to the WEMVA Hessian, $\mathbf{H}_W$). Note from the expansion in Equation B-5 that the forward WEMVA operator constitutes a linear operator that relates perturbations in the background slowness squared to perturbations in the image:

$$\mathbf{\Delta r}_{mig}(\mathbf{s}^2) - \mathbf{\Delta r}_{mig}(\mathbf{s}_0^2) = \mathbf{\Delta}(\mathbf{\Delta r}_{mig}) \approx \mathbf{W\Delta s}^2. \tag{B-6}$$

Now I obtain the gradient of the misfit function (Equation B-4),

$$\begin{aligned}\nabla\Phi(\mathbf{s}^2) &= -\left[\frac{\partial \mathbf{\Delta r}_{mig}(\mathbf{s}^2)}{\partial \mathbf{s}^2}\bigg|_{\mathbf{s}^2=\mathbf{s}_0^2}\right]'[\mathbf{\Delta r}_{mig}(\mathbf{s}_0^2) + \mathbf{W\Delta s}^2] \\ &= -\mathbf{W}'[\mathbf{\Delta r}_{mig}(\mathbf{s}_0^2) + \mathbf{W\Delta s}^2],\end{aligned} \tag{B-7}$$

Finally, I can initialize the conjugate directions method by setting the initial model and the initial residual, the latter evaluating the fitting goal on the initial model (see Appendix A),

$$\mathbf{\Delta s}_0^2 = \mathbf{0}; \quad \mathbf{f}_0 = \mathbf{\Delta r}_{mig}(\mathbf{s}_0^2) + \mathbf{W\Delta s}_0^2 = \mathbf{\Delta r}_{mig}(\mathbf{s}_0^2). \tag{B-8}$$

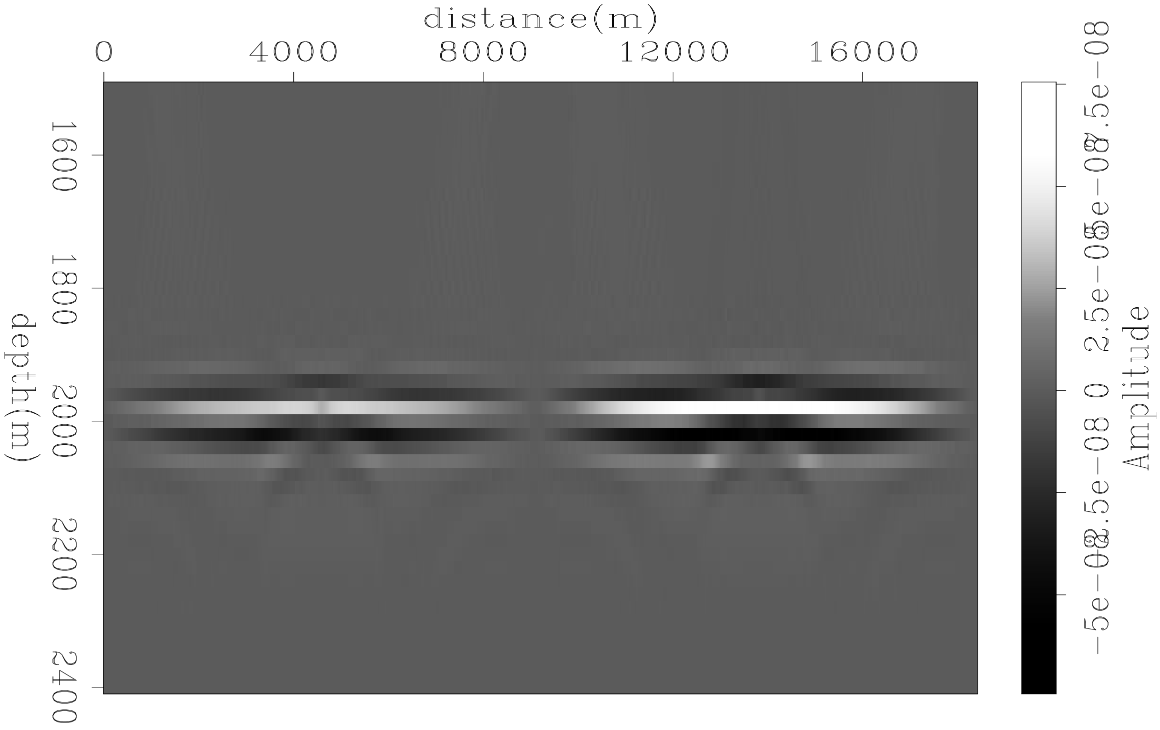


Figure 13: Reflectivity comparison, $\lambda = 1.5$. *Left*: LWI reflectivity; *Right*: LWIVU reflectivity [**CR**]. alejandro1/. Fig-R-1-5

The gradient is defined with the *negative* of adjoint operator (Equation B-7) applied to the residuals,

$$\mathbf{g}_k = -\mathbf{W}'\mathbf{f}_k. \tag{B-9}$$

and it is projected onto the data space with the *negative* of the forward operator

$$\mathbf{G}_k = -\mathbf{W}\mathbf{g}_k. \tag{B-10}$$

Note that the negative signs accompanying the forward and the adjoint WEMVA operators in Equations B-9 and B-10 are also utilized in the WEMVA Hessian for the LWIVU operator (see Appendix A).

REFERENCES

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

Barnier, G. and A. Almomin, 2014, Tutorial on two-way wave equation operators for acoustic, isotropic, constant-density media: SEP-Report, **155**, 23–56.

Biondi, B., 2006, 3-D seismic imaging: Society of Exploration Geophysicists.

Biondi, B. and A. Almomin, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79**, no. 3, WA129–WA140.

Biondi, B., E. Biondi, M. Maharramov, and Y. Ma, 2015, Dissection of the full-waveform inversion hessian: SEP-Report, **160**, 19–38.

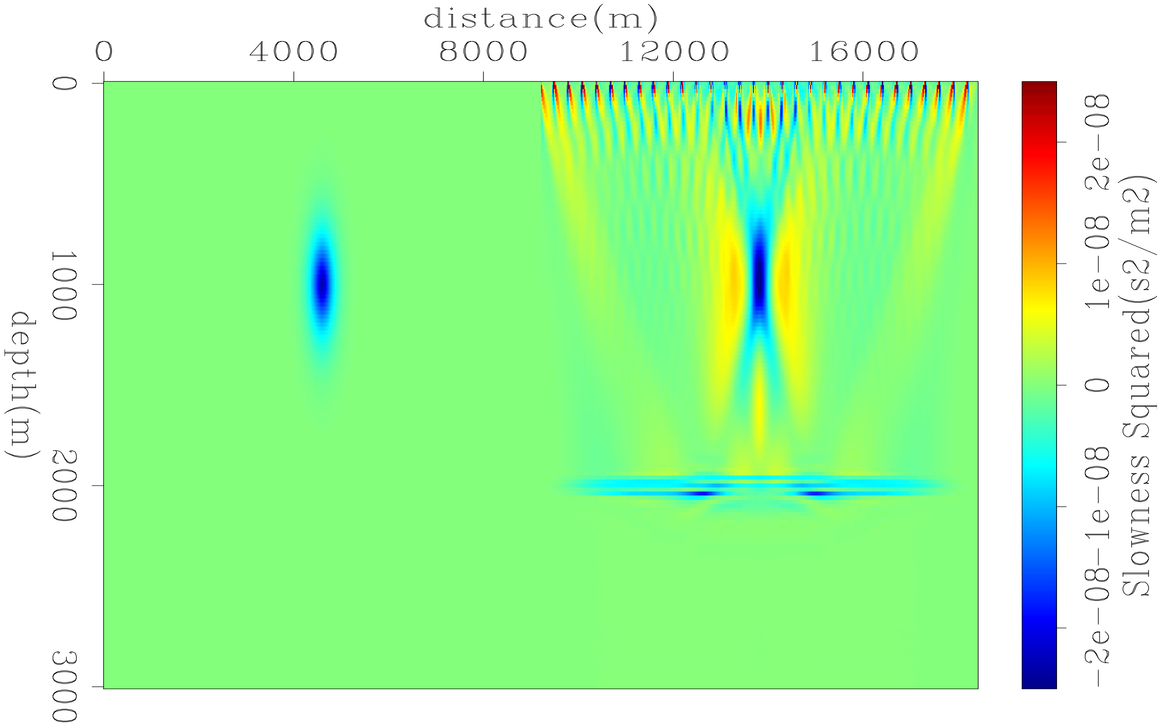


Figure 14: Perturbation in slowness squared, $\lambda = 1.5$. *Left*: True anomaly; *Right*: LWIVU perturbation [**CR**]. alejandro1/. Fig-B-1-5

Cabrales-Vargas, A., B. Biondi, and R. Clapp, 2016, Linearized waveform inversion with (small) velocity updating : SEP-Report, **163**, 189–196.

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

Claerbout, J. F., 1992, Earth soundings analysis: Processing versus inversion: Blackwell Scientific Publications, Inc.

——, 2014, Geophysical image estimation by example: Jon Claerbout.

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

Clapp, R., 2009, Reverse time migration with random boundaries: SEG Technical Program Expanded Abstracts, 2809–2813.

Clapp, R. G., 2010, More fun with random boundaries: SEP-Report, **140**, 95–102.

Dai, W., C. Boonyasiriwat, and G. Schuster, 2010, 3D multi-source least-squares reverse time migration: SEG Technical Program Expanded Abstracts, 3120–3124.

Fichtner, A., 2011, Full seismic waveform modelling and inversion: Springer Science & Business Media.

Fletcher, R., D. Nichols, R. Bloor, and R. Coates, 2016, Least-squares migation - data domain versus image domain using point spread functions: The Leading Edge, **35**, no. 2, 157–162.

Guitton, A., A. Valenciano, D. Bevc, and J. Claerbout, 2007, Smoothing imaging conditions for shot-profile migration: Geophysics, **72**, no. 3, S149–S154.

Hestenes, M. and E. Stiefel, 1952, Methods of conjugate gradients for solving linear systems: Journal of Research of the National Bureau of Standards, **49**, no. 6, 409–436.

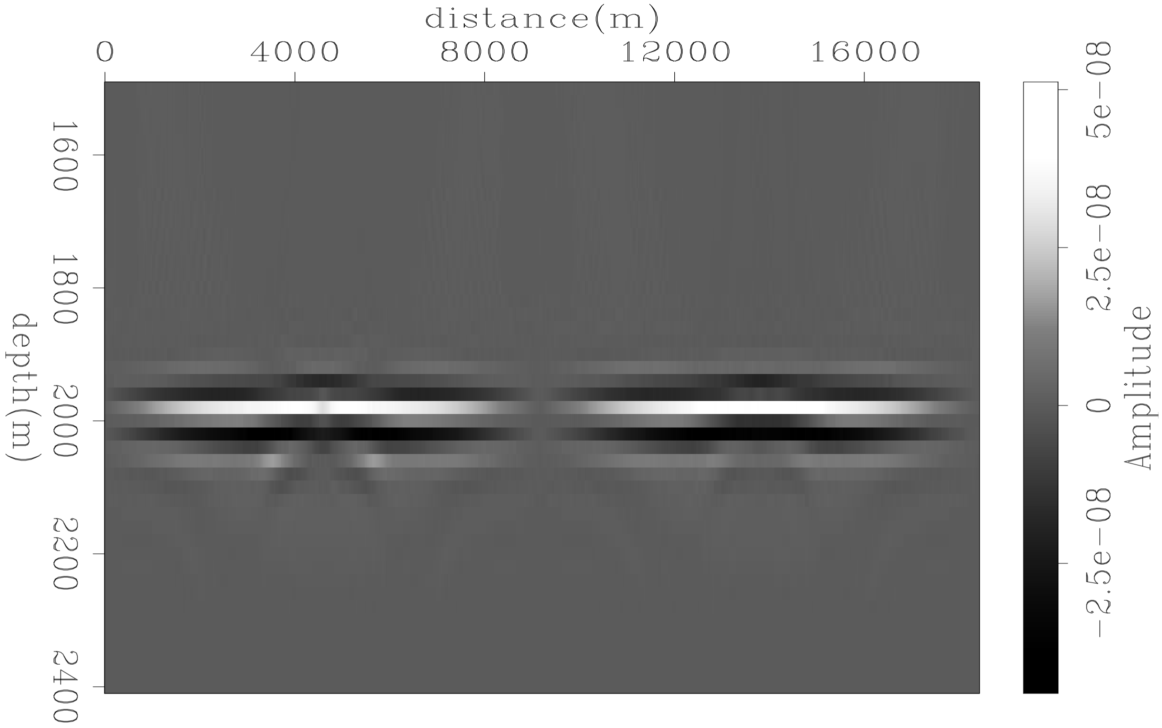


Figure 15: Reflectivity comparison, $\lambda = 2.5$. *Left*: LWI reflectivity; *Right*: LWIVU reflectivity [**CR**]. alejandro1/. Fig-R-2-5

Luo, S. and D. Hale, 2014, Reverse time migration in the presence of velocity errors: Geophysics, **79**, no. 4, S153–S161.

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

Rocha, D., N. Tanushev, and P. Sava, 2016, Acoustic wavefield imaging using the energy norm: Geophysics, **81**, no. 4, S151–S163.

Ronen, S. and C. Liner, 2000, Least-squares DMO and migration: Geophysics, **65**, 1364–1371.

Symes, W. and J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663.

Tang, Y., 2011a, Fast automatic wave-equation migration velocity analysis using encoded simultaneous sources: SEP-Report, **145**, 27–46.

———, 2011b, Imaging and velocity analysis by target-oriented wavefield inversion: PhD thesis, Stanford University.

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

Valenciano, A., B. Biondi, and R. Clapp, 2009, Imaging by target-oriented wave-equation inversion: Geophysics, **74**, no. 6, WCA109–WCA120.

Valenciano, A. A., 2008, Imaging by wave-equation inversion: PhD thesis, Stanford University.

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

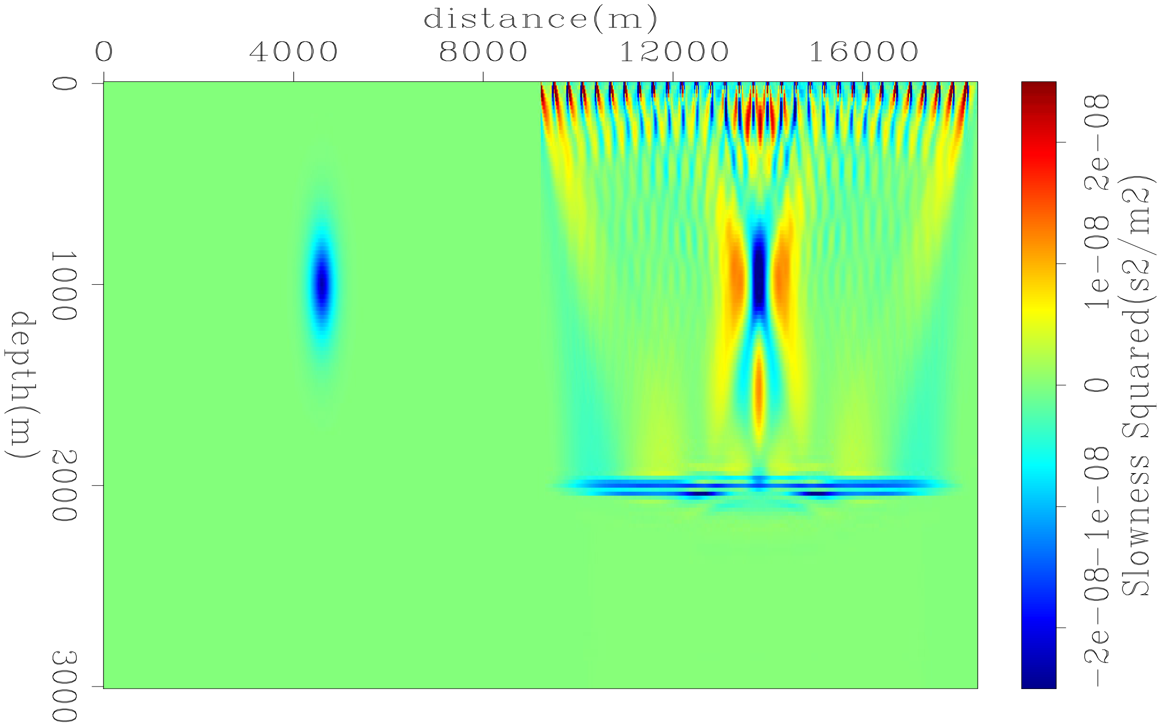


Figure 16: Perturbation in slowness squared, $\lambda = 2.5$. *Left*: True anomaly; *Right*: LWIVU perturbation [**CR**]. alejandro1/. Fig-B-2-5

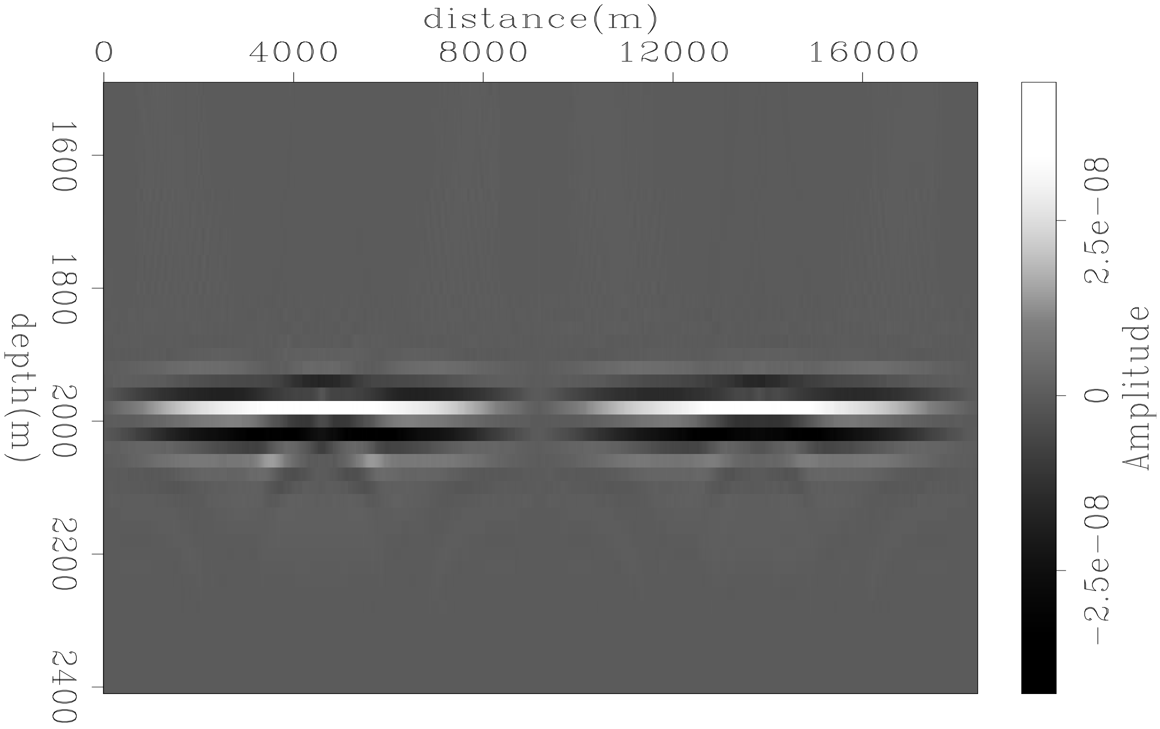


Figure 17: Reflectivity comparison, $\lambda = 2.5$, at 8 iterations. *Left*: LWI reflectivity; *Right*: LWIVU reflectivity [**CR**]. alejandro1/. Fig-R-2-5a

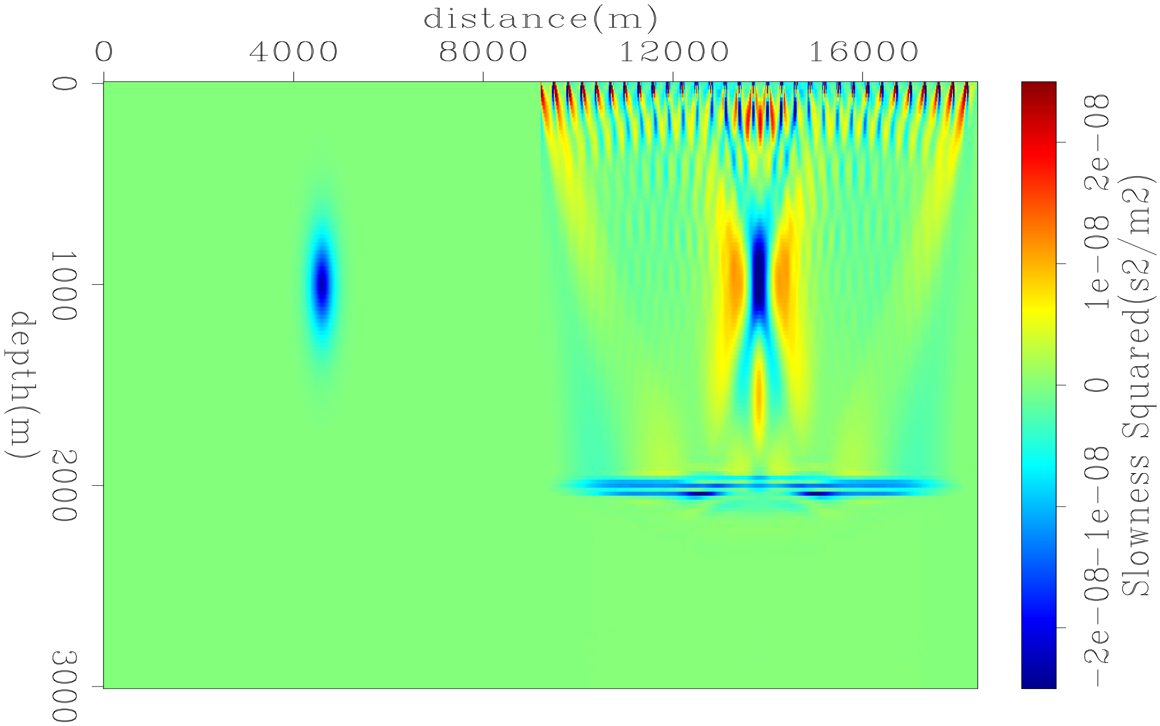


Figure 18: Perturbation in slowness squared, $\lambda = 2.5$, at 8 iterations. *Left*: True anomaly; *Right*: LWIVU perturbation [**CR**]. alejandro1/. Fig-B-2-5a

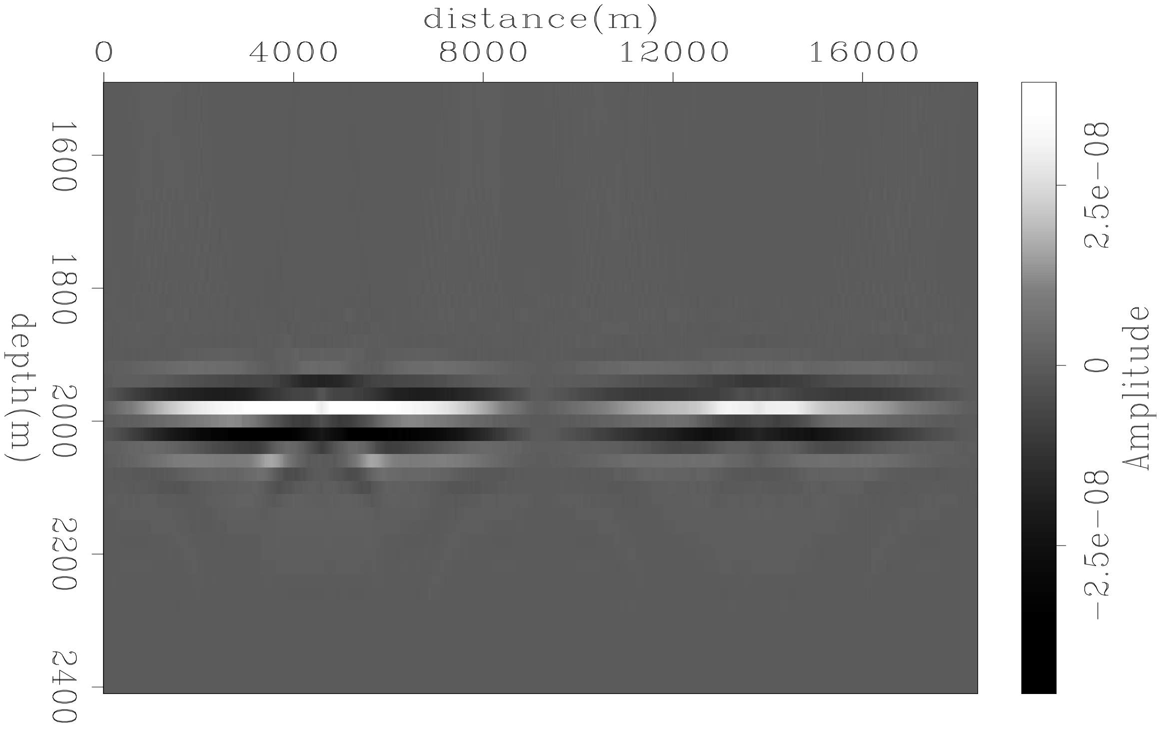


Figure 19: Reflectivity comparison, $\lambda = 5$. *Left*: LWI reflectivity; *Right*: LWIVU reflectivity [**CR**]. alejandro1/. Fig-R-5-0

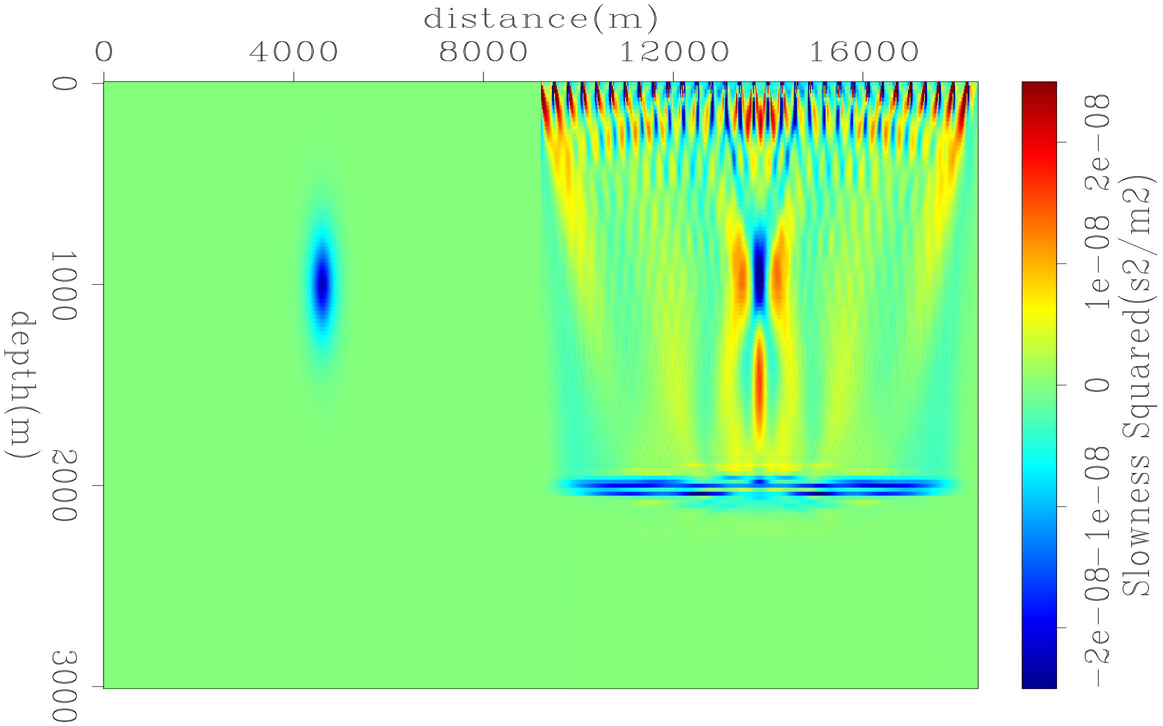


Figure 20: Perturbation in slowness squared, $\lambda = 5$. *Left*: True anomaly; *Right*: LWIVU perturbation [**CR**]. alejandro1/. Fig-B-5-0

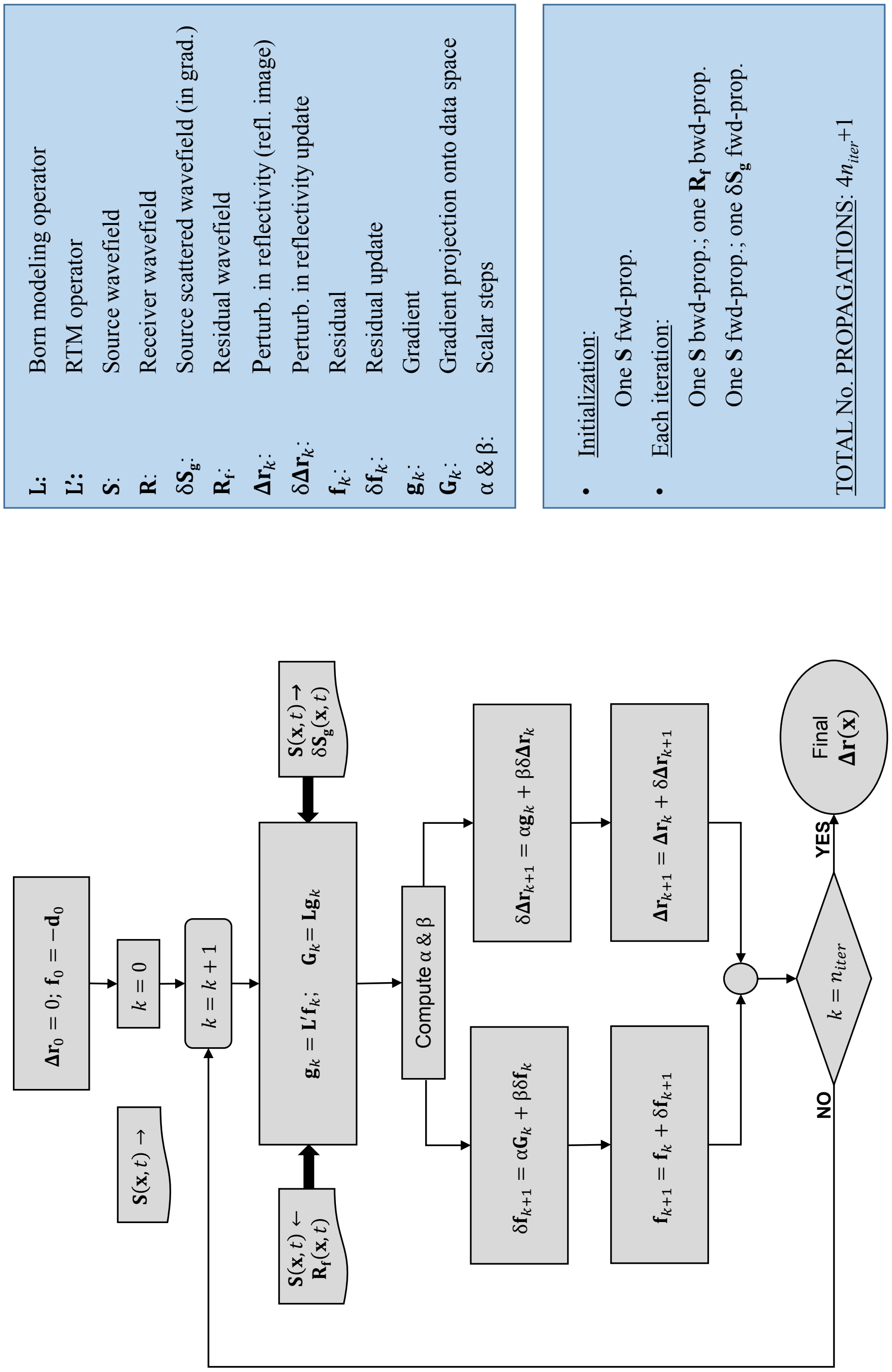


Figure 21: Flowchart of LWI in data space [**NR**]. alejandro1/. FC-LWI

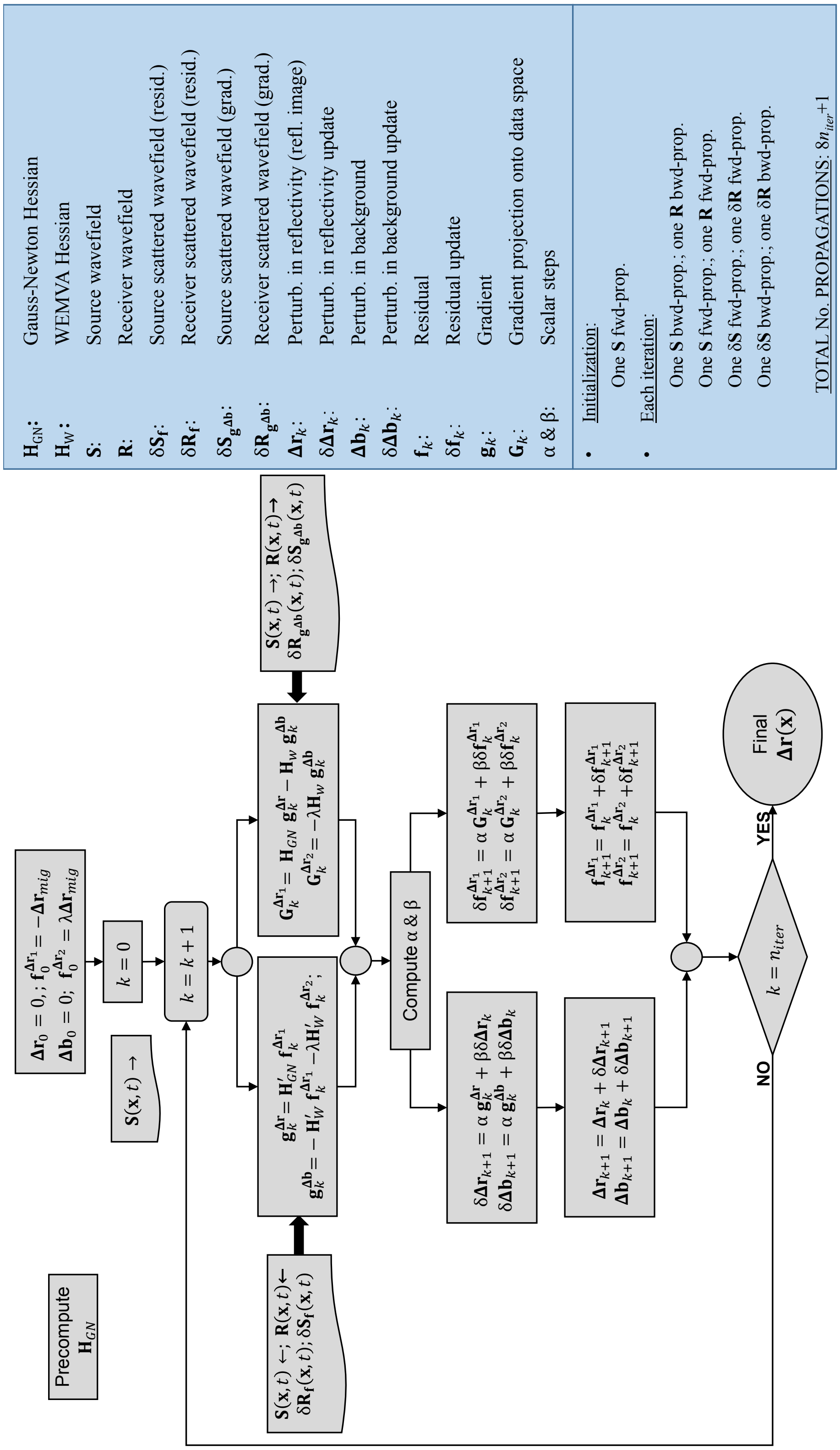


Figure 22: Flowchart of LWIVU [**NR**]. alejandro1/. FC-LWIVU

Solving nonlinear inverse problems by linearized model extension - a survey of possible methods

Biondo Biondi, Rahul Sarkar, and Joseph Jennings

ABSTRACT

Nonlinear inverse problems are challenging because gradient-based inverse algorithms may converge toward local minima instead of the desired global minimum. Different methods can be used to solve nonlinear inverse problems based on a linearized model extension; these methods differ in their global-convergence characteristics and in their convergence rate. We present and analyze, both analytically and numerically, three types of such methods. All three methods show attractive global convergence properties. However, our analysis is both incomplete and based on a simple 1D wave-propagation problem where the medium is characterized by a single slowness value. We discuss the convergence rate of the three types of solution we proposed, but, at the current stage of our research, we cannot reach any definitive conclusions on their convergence rate.

INTRODUCTION

Full waveform inversion (FWI) has a well-known convergence problem when the starting velocity model is far from the correct one and low frequency are not present in the data, or they are too noisy. One of the most promising direction of research for overcoming this problem is based on an extension of the velocity model. The first proposed solutions were based on extension of the reflectivity model (e.g. migrated image); we will refer to all of these methods, somewhat inappropriately, as wave-equation migration velocity analysis (WEMVA) methods (Symes and Carazzone, 1991; Biondi and Sava, 1999; Sava and Biondi, 2004; Shen and Symes, 2008; Zhang and Biondi, 2013). More recently, Symes (2008) and Biondi and Almomin (2014) have proposed extensions of the whole velocity model; that is, of long wavelength as well as short wavelength. These methods have been successful to converge to good models when applied to field data; however, their convergence might be too slow for being directly applicable to large scale problems (Almomin and Biondi, 2014). With another rough, but convenient, generalization we will refer two these methods as tomographic waveform inversion (TFWI) methods.

The main goals of our project are: 1) to develop faster-converging algorithms to apply to the extended FWI methods (e.g. TFWI), and 2) to explore the applicability of the idea of a linearized extension to the solution of other important nonlinear inverse problems in geophysics. In the first section of this report we formalize the idea of applying a linearized model extension to a generic non-quadratic optimization problem.

The second section presents the modeling equation of a simple 1D waveform inversion problem. We assume the data to be a single trace recorded from a transmission experiment between one source and one receiver in an homogeneous medium, with the slowness of the medium being the only parameter to estimate. We also introduce a useful linear extension of the nonlinear modeling operator that will be used to test inverse methods.

In the last sections of this report we present and analyze three ways of solving the nonlinear inverse problem by using a linearized model extension. The first way is based on the optimization of a two-term objective function, and it is related to the TFWI methods proposed by Symes (2008) and Biondi and Almomin (2014); The first term measures data fitting whereas the second one is a regularization (model focusing) term.

The second approach is based on alternating between the optimization of two objective functions. It is conceptually related to WEMVA methods as presented by Biondi and Sava (1999); Sava and Biondi (2004); Zhang and Biondi (2013) (e.g. migration and velocity model updating), where the velocity updating is driven by the matching of the current image to a better focused image.

The third proposed methods minimizes an objective function with a single term. This single term depends on the slowness model through two modeling operators: the original nonlinear operator and the extended linear operator. Our analysis shows that this one-term objective function has some characteristics of both the FWI and the WEMVA objective functions; we will refer to it as the FWI-WEMVA objective function. A version of this objective function was first presented by Symes (2008). In that paper this objective function was used as the basis for the formalization of the velocity analysis problem as a constrained optimization problem, rather than to be directly minimized.

SOLVING NON-QUADRATIC OPTIMIZATION PROBLEMS BY LINEARIZED MODEL EXTENSION

As mentioned above, one of our goals is to explore the applicability of the idea of a linearized extension to the solution of other important nonlinear inverse problems in seismology. Therefore, we first formalize the inversion problem in general terms, and then we study several inversion approaches by analyzing their application to a specific 1D waveform inversion problem.

We want to estimate the vector of model parameters, $\mathbf{m}$, from the recorded data vector, $\mathbf{d}_r$, recorded as the output of an operator $\mathcal{L}$ that is non-linear with respect to the model parameters; that is,

$$\mathbf{d}_r = \mathcal{L}\left(\overline{\mathbf{m}}\right), \tag{1}$$

where $\overline{\mathbf{m}}$ is the "true" value of the parameter vector, and it is the ideal solution of the estimation problem. We can set up the estimation as the least-squares problem of minimizing

$$J\left(\mathbf{m}\right) = \frac{1}{2}\left\|\mathcal{L}\left(\mathbf{m}\right) - \mathbf{d}_r\right\|_2^2. \tag{2}$$

Because of the non-linear dependency between the modeled data, $\mathbf{d} = \mathcal{L}\left(\mathbf{m}\right)$ and the parameter vector $\mathbf{m}$, the objective function in equation 2 is not quadratic, and in general, is not even convex, and presents many local minima. Therefore, when we apply gradient-based methods to solve the optimization problem in 2 we are likely to converge towards a local minimum, instead of the desired global one ($\overline{\mathbf{m}}$).

We are interested in improving the convergence towards the global minimum by solving a different optimization problem that shares the global minimum with the one expressed in 2, but does not have local minima, or at least, can be “safely driven” to converge to the global minimum starting from an arbitrary starting solution $\mathbf{m}_0$. We start by extending the non-linear operator by adding to the modeled data ($\mathbf{d}$) the output, ${}_{\mathrm{L}}\mathbf{d}$, of an appropriately defined linear operator, $\tilde{\mathbf{L}}$, applied to an additional parameter vector ${}_{\mathrm{L}}\mathbf{m}$. The“total” modeled data vector is thus expressed as,

$$ {}_{\mathrm{T}}\mathbf{d} = \tilde{\mathcal{L}}\left(\mathbf{m}, {}_{\mathrm{L}}\mathbf{m}\right) = \mathcal{L}\left(\mathbf{m}\right) + \tilde{\mathbf{L}}\left(\mathbf{m}\right) {}_{\mathrm{L}}\mathbf{m} = \mathbf{d} + {}_{\mathrm{L}}\mathbf{d}, \tag{3} $$

where $\tilde{\mathcal{L}}$ is the “extended” modeling operator that is function of both the original parameter vector, $\mathbf{m}$, and the *extended model* parameter vector, ${}_{\mathrm{L}}\mathbf{m}$. Notice that the linear operator $\tilde{\mathbf{L}}$ is itself a non-linear function of the parameter vector $\mathbf{m}$.

The optimization problem in 2 can be modified into the following

$$ J_{\mathrm{E}}\left(\mathbf{m}, {}_{\mathrm{L}}\mathbf{m}\right) = \frac{1}{2}\left\|\tilde{\mathcal{L}}\left(\mathbf{m}, {}_{\mathrm{L}}\mathbf{m}\right) - \mathbf{d}_r\right\|_2^2, \tag{4} $$

that obviously has a global minimum for ($\mathbf{m} = \overline{\mathbf{m}}, {}_{\mathrm{L}}\mathbf{m} = 0$). In a more compact notation, if we combine the two model vectors into one, we can write

$$ J_{\mathrm{E}}\left({}_{\mathrm{T}}\mathbf{m}\right) = \frac{1}{2}\left\|\tilde{\mathcal{L}}\left({}_{\mathrm{T}}\mathbf{m}\right) - \mathbf{d}_r\right\|_2^2, \tag{5} $$

where ${}_{\mathrm{T}}\mathbf{m} = (\mathbf{m}, {}_{\mathrm{L}}\mathbf{m})$ and $\overline{{}_{\mathrm{T}}\mathbf{m}} = (\mathbf{m} = \overline{\mathbf{m}}, {}_{\mathrm{L}}\mathbf{m} = 0)$.

The extension operator $\tilde{\mathbf{L}}$ plays an important role in the method, and it should be defined according to the specific non-linear behavior of $\mathcal{L}\left(\mathbf{m}\right)$ that prevents convergence to the global minimum of J in the practical problems that we want to address. Ideally, $\frac{d\tilde{\mathcal{L}}}{d_{\mathrm{T}}\mathbf{m}}$ is close to be a unitary operator, or at least the following approximation is valid:

$$ \begin{aligned} \frac{d\tilde{\mathcal{L}}}{d_{\mathrm{T}}\mathbf{m}}\left(\mathbf{m}\right)\frac{d\tilde{\mathcal{L}}}{d_{\mathrm{T}}\mathbf{m}}^{\prime}\left(\mathbf{m}\right)\left[\mathcal{L}\left(\mathbf{m}\right) - \mathbf{d}_r\right] &= \\ \frac{d\tilde{\mathcal{L}}}{d_{\mathrm{T}}\mathbf{m}}\left(\mathbf{m}\right)\frac{d\tilde{\mathcal{L}}}{d_{\mathrm{T}}\mathbf{m}}^{\prime}\left(\mathbf{m}\right)\left[\mathcal{L}\left(\mathbf{m}\right) - \mathcal{L}\left(\overline{\mathbf{m}}\right)\right] &\approx \left[\mathcal{L}\left(\mathbf{m}\right) - \mathcal{L}\left(\overline{\mathbf{m}}\right)\right] \ \forall\, \overline{\mathbf{m}} \text{ and } \mathbf{m}. \end{aligned} \tag{6} $$

In practice, there is not a single linear extension that fulfills the requirement expressed in 6 for all possible events present in typical reflection-seismic data, and for any possible error in starting velocity model. For example, Biondi and Almomin (2014) showed that a time-lag extension of the slowness model is capable of modeling the kinematics of the data residuals caused by long-wavelength errors in the velocity model. However, the same linear extension is not as effective when the nonlinear wave phenomena that hamper global convergence are related to multiple-scattering by discrete interfaces; e.g. multiples.

Focusing operator

The optimization problem in 4 (and 5) is under-constrained. For any choice of the extended model vector $\mathbf{m}$, it is likely that there is a corresponding value of ${}_{\mathrm{L}}\mathbf{m}$ that minimizes J_{E}. Furthermore, the extension of the modeling operator is only an end to achieve the goal of robustly converging toward $\overline{\mathbf{m}}$. At convergence, or even in a neighborhood of the global minimum, we would would like to have the contributions of the linearized extension to be negligible. To achieve this goal, we introduce a “focusing” $\mathcal{F}$ operator.

An essential property of the focusing operator is that its output is more focused than its input. If we define an optimally focused model vector $\overline{{}_{\mathrm{L}}\mathbf{m}}$ when all the energy is focused in at one particular model coordinate $\overline{\mathbf{x}_f}$, we can define a measure of the defocusing as

$$D\left({}_{\mathrm{L}}\mathbf{m}\right) = \left\|\mathbf{X}_f\,{}_{\mathrm{L}}\mathbf{m}\right\|_2^2, \tag{7}$$

where $\mathbf{X}_f$ is a diagonal matrix with the absolute value of the distance from $\overline{\mathbf{x}_f}$; that is, $\mathbf{X}_f = \mathbf{diag}(|\mathbf{x_f} - \overline{\mathbf{x_f}}|)$. By construction $D\left(\overline{{}_{\mathrm{L}}\mathbf{m}}\right) = 0$.

The focusing operator must have at least the property that

$$\begin{cases} \mathcal{F}\left({}_{\mathrm{L}}\mathbf{m}\right) = \overline{{}_{\mathrm{L}}\mathbf{m}} & \text{if } {}_{\mathrm{L}}\mathbf{m} = \overline{{}_{\mathrm{L}}\mathbf{m}} \\ D\left(\mathcal{F}\left({}_{\mathrm{L}}\mathbf{m}\right)\right) < D\left({}_{\mathrm{L}}\mathbf{m}\right) & \text{if } {}_{\mathrm{L}}\mathbf{m} \neq \overline{{}_{\mathrm{L}}\mathbf{m}}. \end{cases} \tag{8}$$

In the following sections we discuss how the focusing operator can be introduced in a regularization term to be added the data-fitting term of the objective function, or how we can directly introduce it into the data-fitting term of the objective function. Depending on the way that the focusing operator is introduced in the objective function(s), and on the choice of $\tilde{\mathbf{L}}$, other properties, in addition to 8, of $\mathcal{F}$ are required for defining an estimation method that robustly converges towards the global minimum.

SIMPLE 1D WAVEFORM INVERSION PROBLEM

In the following sections we present several approaches to solve the estimation problem. We analyze the behavior of these methods using a simple 1D "wavelet-shift" modeling operator. In a homogeneous medium with slowness $\bar{s}$, the wavefield generated by one source and recorded by one receiver at a distance l is equal, when we ignore an amplitude scaling, to the source wavelet time-shifted by $\Delta t = ls$. The recorded-data vector $\mathbf{d}_r \in \Re^{N_t}$ is expressed as follows:

$$\mathbf{d}_r = \mathcal{L}\left(\bar{s}\right) = \mathbf{S}\left(l\bar{s}\right)\mathbf{g}, \tag{9}$$

where $\mathbf{g} \in \Re^{N_g}$ is the source vector, and $\mathbf{S}\left(\Delta t\right) \in \Re^{N_t \mathrm{x} N_g}$ is a time-shift operator that shifts an input vector by the time interval Δt.

We define the linearized extension of modeling as the time convolution of the shifted wavelet with a filter c; that is as,

$${}_{\mathrm{L}}d\left(t\right) = g\left(t - ls\right) *_{\tau} c\left(\tau\right), \tag{10}$$

where with the symbol $*_{\tau}$ we denote convolution with the filter $c\left(\tau\right)$ along the time-lag axis τ. For matter of convenience, we fix the value of c at the origin to be equal to zero; that is, $c\left(\tau = 0\right) = 0$.

The data vector, ${}_{\mathrm{L}}\mathbf{d}$, produced by the linearized extension can be expressed by the following matrix-vector product

$${}_{\mathrm{L}}\mathbf{d} = \tilde{\mathbf{L}}\left(s\right)\mathbf{c}, \tag{11}$$

where $\tilde{\mathbf{L}}\left(s\right) \in \Re^{N_t \mathrm{x} N_\tau}$ is a matrix appropriately constructed with the elements of the shifted source vector $\mathbf{S}\left(l\bar{s}\right)\mathbf{g}$, and $\mathbf{c}$ is a vector of length N_τ representing the discretization of convolutional filter c. The total modeled data ${}_{\mathrm{T}}\mathbf{d}$ can thus be expressed as follows:

$${}_{\mathrm{T}}\mathbf{d} = \tilde{\mathcal{L}}\left(s, \mathbf{c}\right) = \mathcal{L}\left(s\right) + \tilde{\mathbf{L}}\left(s\right)\mathbf{c} = \mathbf{S}\left(ls\right)\mathbf{g} + \tilde{\mathbf{L}}\left(s\right)\mathbf{c}. \tag{12}$$

As mentioned before, we set the zero-lag coefficient of $\mathbf{c}$ to zero; that is, $c_0 = 0$. For convenience, we also assume that $\mathbf{c}$ is centered around the origin; that is, $c_i \neq 0$ for $N_\tau/2 \leq i \leq N_\tau/2$ with N_τ even.

Notice that our choice of $\tilde{\mathbf{L}}$ is, in purpose, not the most obvious one. The most natural choice would have been to define $\tilde{\mathbf{L}}\mathbf{c}$ as the convolution of $\mathbf{c}$ with the first time derivative of the shifted source wavelet; that is, to build $\tilde{\mathbf{L}}$ with elements of $-l\mathbf{S}\left(l\bar{s}\right)\dot{\mathbf{g}}$, instead of $\mathbf{S}\left(l\bar{s}\right)\mathbf{g}$ (Biondi and Almomin, 2014). In practice, there is a little differences between these two choices. The main reason we made this choice is to emphasize that the definition of $\tilde{\mathbf{L}}$ does not need to be based on physical arguments. The main criterion driving the definition of $\tilde{\mathbf{L}}$ should be to fulfill the approximation expressed in 6 as accurately as possible.

Linearization of modeling operators

When computing gradients of the proposed objective functions, we need to linearize the modeling operators with respect to perturbations of the model parameters. In this section we derive the linearization of the simple 1D modeling operator $\tilde{\mathcal{L}}\left(s,\mathbf{c}\right)$ with respect to slowness s and the coefficients of the convolutional filter $\mathbf{c}$. Taking advantage of the fact that slowness is a scalar, the operators derived from these linearizations can also be expressed as vector and matrices.

The derivative of the operator $\mathcal{L}\left(s\right)$ with respect to slowness, that we will denote as the vector $\mathbf{L} \in \Re^{N_t}$, is given by the following:

$$\frac{\partial \mathcal{L}\left(s\right)}{\partial s} = \mathbf{L}\left(s\right) = -l\mathbf{S}\left(ls\right)\dot{\mathbf{g}}, \tag{13}$$

where $\dot{\mathbf{g}}$ is the discretization of the first time derivative of the source function $g\left(t\right)$.

The derivative of the extension operator $\tilde{\mathbf{L}}\left(s,\mathbf{c}\right)$ with respect to slowness is given by the matrix-vector product of the matrix $\dot{\tilde{\mathbf{L}}} \in \Re^{N_t \mathrm{x} N_\tau}$ with the vector $\mathbf{c}$ as follows:

$$\frac{\partial \tilde{\mathbf{L}}}{\partial s}\left(s,\mathbf{c}\right) = \dot{\tilde{\mathbf{L}}}\left(s\right)\mathbf{c}. \tag{14}$$

As for the matrix $\tilde{\mathbf{L}}$ (equation 11), also the matrix $\dot{\tilde{\mathbf{L}}}$ can be appropriately constructed with the elements of the first time derivative of the shifted source vector $\mathbf{S}\left(ls\right)\dot{\mathbf{g}}$ scaled by $-l$.

Since ${}_{\mathrm{L}}\mathbf{d}$ is linear with respect to the convolutional filter $\mathbf{c}$, the linearization of the extension operator $\tilde{\mathbf{L}}\left(s,\mathbf{c}\right)$ with respect to $\mathbf{c}$ is simply the matrix $\tilde{\mathbf{L}}$; that is,

$$\frac{\partial \tilde{\mathbf{L}}}{\partial \mathbf{c}}\left(s,\mathbf{c}\right) = \tilde{\mathbf{L}}. \tag{15}$$

Numerical example

We illustrate the properties of the operators defined above by using a numerical example. Figure 1 shows the results of this numerical example.

Figure 1a shows the wavelet that we used for this example. It was derived by taking the first time derivative of a Ricker wavelet with fundamental frequency of 7 Hz. For convenience, in the proceedings we will refer to it as the *Ricker-derived wavelet.* Figure 1b shows the data residuals ($\mathcal{L}(s_o) - \mathcal{L}(\bar{s})$) for a range of starting slowness values (s_o), and a true slowness $\bar{s} = 1$ s/km. The source-receiver distance l is 4 km.

Figures 1c and 1d show the *back projection* of the data residuals into the model space. These back projections are an important component of any gradient-based estimation algorithm. Figure 1c shows the application of the adjoint of $\mathbf{L}(s_o)$ to the corresponding data residuals shown in Figure 1b. Similarly, Figure 1d shows the application of the adjoint of $\tilde{\mathbf{L}}(s_o)$ to the same data residuals shown in Figure 1b.

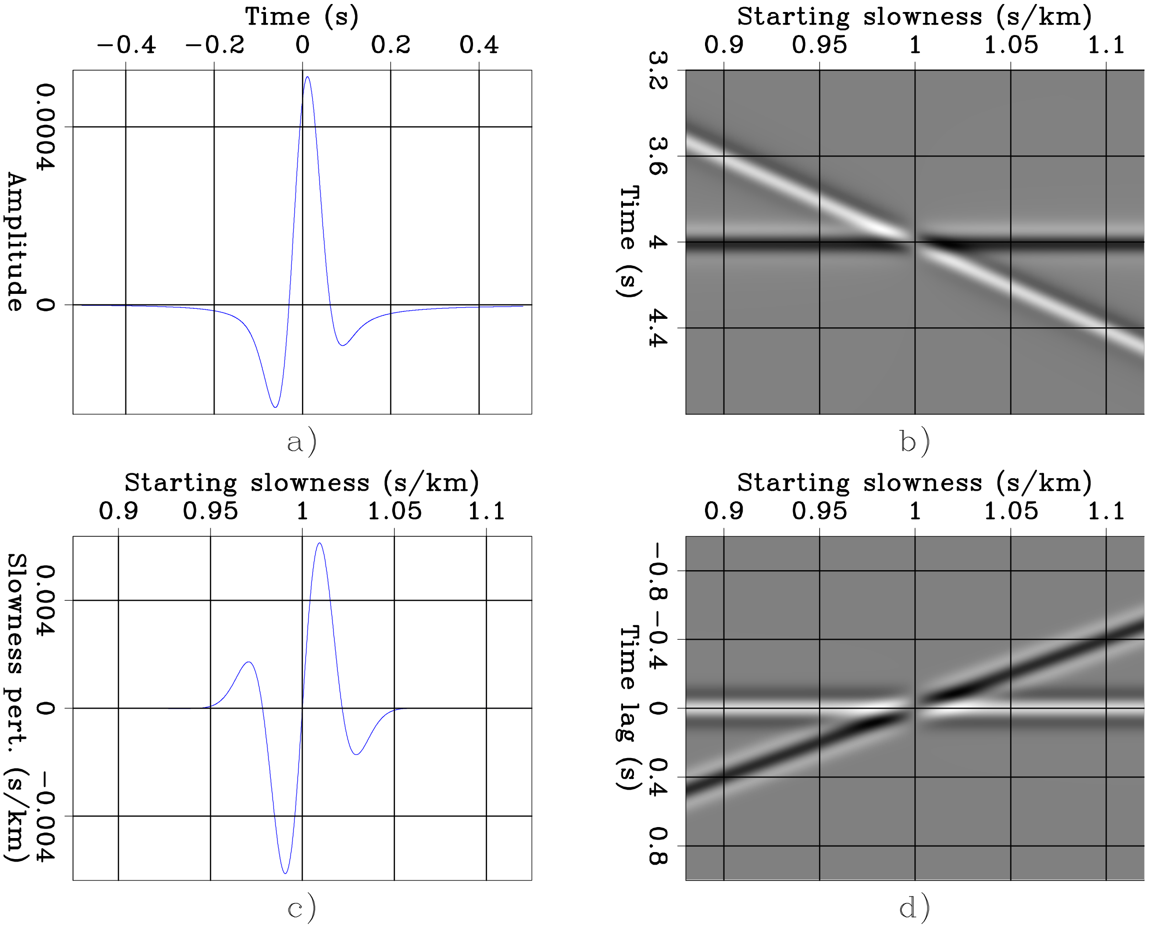
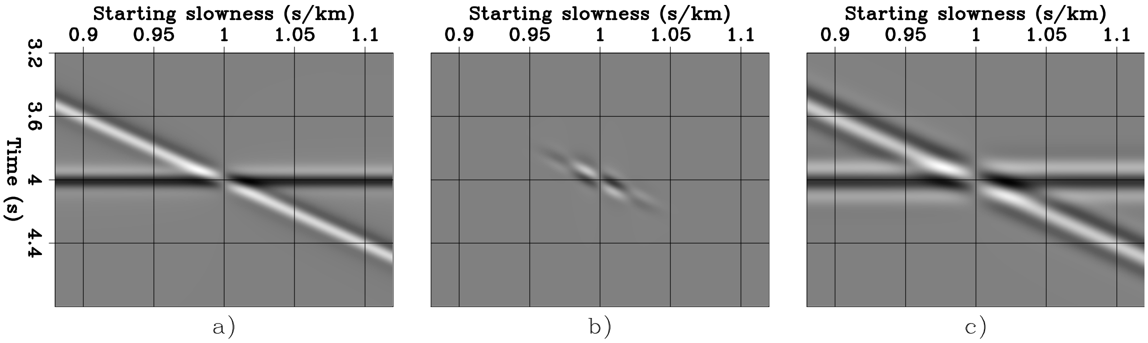


Figure 1: a) Ricker-derived wavelet used for the numerical examples. b) data residuals for a range of starting slowness values. c) back projection of the data residuals into slowness space for the same range of starting slowness values. d) back projection of the data residuals into the convolutional filter space. [**ER**] biondo1/. Examp-1D

As discussed, above, an important characteristic of the extended modeling operator is that its linearization should be close to be unitary, in the sense defined by the relation in equation 6. Figure 2 shows that the original (non extended) linearized operator is far from fulfilling that condition, whereas the linearized extension fulfills that condition, at least with regards to the kinematics of the reconstructed residuals.

Figure 2b shows the application of $\mathbf{LL}'$ to the data residuals shown in Figure 1b, which for convenience of the reader are also displayed in Figure 2a. Figure 2b can also be described as the *forward projection* into the data space of the slowness perturbations shown in Figure 1c. The data residuals are well reconstructed only in a small interval of starting slowness centered around the true slowness ($\bar{s}$). Furthermore, at both ends of this range, the reconstructed residuals have the wrong polarity. In contrast, Figure 2c shows the application of $\tilde{\mathbf{L}}\tilde{\mathbf{L}}'$ to the data residuals. It can also be described as the forward projection into the data space of the convolutional filter perturbations shown in Figure 1d. The kinematics of the events in the data residuals are well reconstructed. The only noticeable difference between panel 2a and panel 2c is that the reconstructed events are convolved with the square of the source wavelet.



a) b) c)

Figure 2: a) Data residuals $[\mathcal{L}(s_o) - \mathcal{L}(\bar{s})]$. b) $\mathbf{LL}'[\mathcal{L}(s_o) - \mathcal{L}(\bar{s})]$. c) $\tilde{\mathbf{L}}\tilde{\mathbf{L}}'[\mathcal{L}(s_o) - \mathcal{L}(\bar{s})]$. [**ER**] biondo1/. Data-Examp-1D

Focusing operators

We tested different focusing operators; all of them are linear with respect to the convolutional filter $\mathbf{c}$. Therefore, we will denote as $\mathbf{F}$ and write $\mathcal{F}(\mathbf{c}) = \mathbf{Fc}$. There are two families of focusing operators that can be useful. The operators belonging to the first family scale the filter coefficients as a function of the distance from the origin; that is, as a function of the time lag τ. We refer to the focusing operators belonging to this family as *amplitude* focusing operators because they simply scale the amplitudes of the input filter. The operators belonging to the second family "shift" the filter coefficients towards the origin. Therefore, we refer to these operators as *phase* focusing operators because they actually change the phase of the convolutional filter.

DSO (Differential Semblance Operator)

The first operator we analyze is derived from the classical DSO operator (Symes and Carazzone, 1991) and defined as $\mathbf{F}_D = \mathbf{diag}(\tau_f)$ with

$$\tau_f = \begin{cases} \frac{\tau_{\rm w}\tau_{\max} - |\tau|}{\tau_{\rm w}\tau_{\max}} & \text{if } |\tau| < \tau_{\rm w}\tau_{\max} \\ 0 & \text{if } |\tau| \geq \tau_{\rm w}\tau_{\max}, \end{cases} \tag{16}$$

where $\tau_{\max}$ is the maximum τ represented in $\mathbf{c}$, and $\tau_{\rm w}$ (with $0 \leq \tau_{\rm w} \leq 1$) is an adimensional parameter that determines the width of the triangular window; this parameter may change with iterations. Notice that the classical DSO operator is equal to $\mathbf{I} - \mathbf{F}_D$ with $\tau_{\mathrm{w}} = 1$.

Gaussian window

The second operator we analyze is also a simple amplitude focusing operator; it is defined as the truncated Gaussian window $\mathbf{F}_G = \mathbf{diag}(\tau_f)$, with

$$\tau_f = \begin{cases} e^{-5\frac{\tau^2}{(\tau_{\mathrm{W}}\tau_{\max})^2}} & \text{if } |\tau| < \tau_{\mathrm{W}}\tau_{\max} \\ 0 & \text{if } |\tau| \geq \tau_{\mathrm{W}}\tau_{\max}, \end{cases} \tag{17}$$

where $\tau_{\max}$ and τ_{W} have the same meaning as in equation 16.

Shift

The third operator we analyze is the simplest phase focusing operator (Almomin, personal communication). The operator $\mathbf{F}_S$ shifts by one sample the coefficients of $\mathbf{c}$ towards the origin. The i-th coefficient $\hat{c}_i$ of the output filter are computed from the coefficients c_i of the input filter as:

$$\hat{c}_i = \begin{cases} c_{i+1} & \text{if } i > 0 \\ c_{i-1} & \text{if } i < 0. \\ 0 & \text{if } |i| = N_\tau/2 \text{ or } i = 0. \end{cases} \tag{18}$$

Shrink

The fourth, and last, operator we analyze is also a phase focusing operator; it scales the τ axis of its input filter by a factor α, that is $\hat{c}(\tau) = c(\alpha\tau)$, with $\alpha \geq 1$. The operator $\mathbf{F}_\alpha$ is the discrete implementation of this axis-shrinking operator that employs a sinc interpolator.

Examples of application of focusing operators

Figure 3 shows the outputs of these four focusing operators as applied to the convolutional filters shown in Figure 1d. Figures 3a, 3b, 3c, and 3d show the results of applying $\mathbf{F}_D$, $\mathbf{F}_G$, $\mathbf{F}_S$, and $\mathbf{F}_\alpha$, respectively. The output of $\mathbf{F}_S$ is indistinguishable from its input because $\mathbf{F}_S$ shifts its input by one sample only.

Figure 4 shows the plots of the defocusing measure D defined in equation 7 to the input and output $\mathbf{c}$ vectors shown in Figure 1d and Figure 3. It demonstrates with a numerical example that all four focusing operators fulfill the condition introduced in 8. For this example, we set $\tau_{\mathrm{W}} = 1.0$ for both $\mathbf{F}_D$ and $\mathbf{F}_G$, and $\alpha = 1.111$ for $\mathbf{F}_\alpha$. The functions are obviously symmetric around $s_o = 1$ s/km, and thus for clarity we plotted them only for $s_o \geq 1$ s/km. The values of $D(\mathbf{F}_S\mathbf{c})$ are uniformly smaller than the values of $D(\mathbf{c})$; however, their respective plots shown in Figure 4 almost perfectly overlap because their difference is tiny.

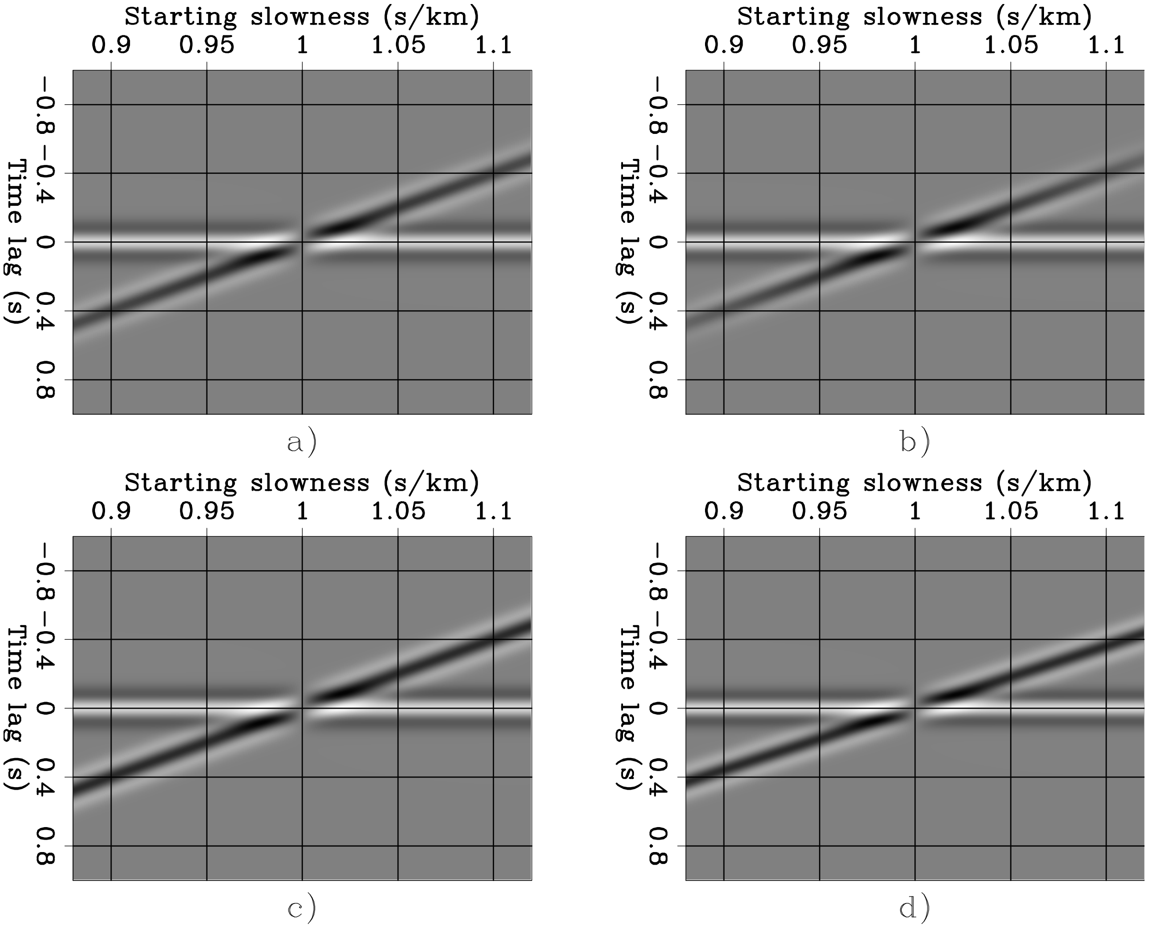


Figure 3: Results of applying the four focusing operators to the convolutional filters $\mathbf{c}$ shown in Figure 1d: a) $\mathbf{F}_D\mathbf{c}$, b) $\mathbf{F}_G\mathbf{c}$, c) $\mathbf{F}_S\mathbf{c}$, and d) $\mathbf{F}_\alpha\mathbf{c}$. For both $\mathbf{F}_D$ and $\mathbf{F}_G$, $\tau_{\rm W} = 1.$, and $\alpha = 1.111$ for $\mathbf{F}_\alpha$. **[ER]** biondo1/. Focus-1D

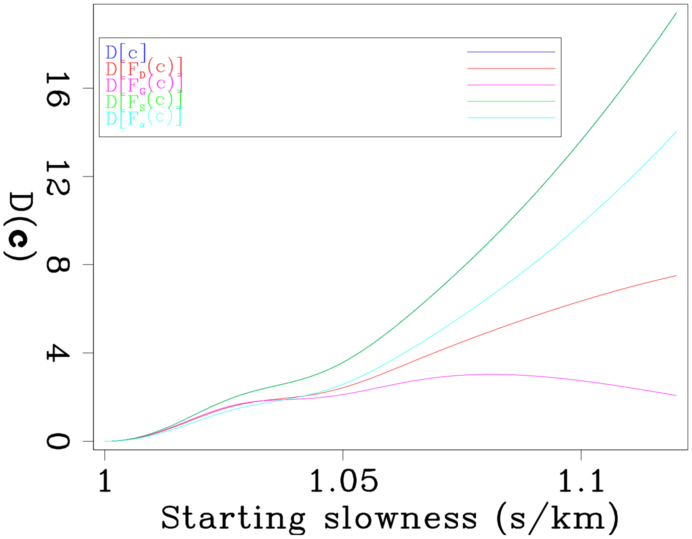


Figure 4: Plots of: $D(\mathbf{c})$, $D(\mathbf{F}_D\mathbf{c})$, $D(\mathbf{F}_G\mathbf{c})$, $D(\mathbf{F}_S\mathbf{c})$, and $D(\mathbf{F}_\alpha\mathbf{c})$. For both $\mathbf{F}_D$ and $\mathbf{F}_G$, $\tau_{\rm W} = 1.$, and $\alpha = 1.111$ for $\mathbf{F}_\alpha$. Notice that the plot of $D(\mathbf{F}_S\mathbf{c})$ almost perfectly overlaps the plot of $D(\mathbf{c})$; however, it is below it for all values of s_o. **[ER]** biondo1/. FM-Focus-1D

TFWI OBJECTIVE FUNCTIONS

Model-space regularization

As previously discussed, there are three different ways in which we constrain problem 4 using the focusing operator $\mathcal{F}$. One of these approaches is to use it in a model regularization term in addition to the data-fitting term defined in problem 4. For the simple 1D problem described above, we can write this objective function as:

$$J_{\mathrm{M}}(s, \mathbf{c}) = \frac{1}{2}\left\|\tilde{\mathcal{L}}(s, \mathbf{c}) - \mathbf{d}_r\right\|_2^2 + \frac{\epsilon}{2}\left\|(\mathbf{I} - \mathbf{F})\,\mathbf{c}\right\|_2^2, \tag{19}$$

where $\mathbf{F}$ is the linear focusing operator. Through our numerical experiments, we have found that choosing $\mathbf{F}$ as the DSO operator provides the best convergence properties. This observation is related to the fact that the phase-focusing operators are nearly unitary and therefore, the result of their forward and adjoint applications in the computation of the gradients does not provide the correct focusing of the extended model space.

The results of running 7000 iterations of a non-linear conjugate gradient inversion are shown in Figure 5. For this inversion, the starting physical slowness (s_0) was 1.12 s/km and the starting extended slowness ($\mathbf{c}_0$) was $\mathbf{0}$. For the regularization, we set $\epsilon = 100$ and $\mathbf{F}$ to be the DSO operator.

Data-space regularization

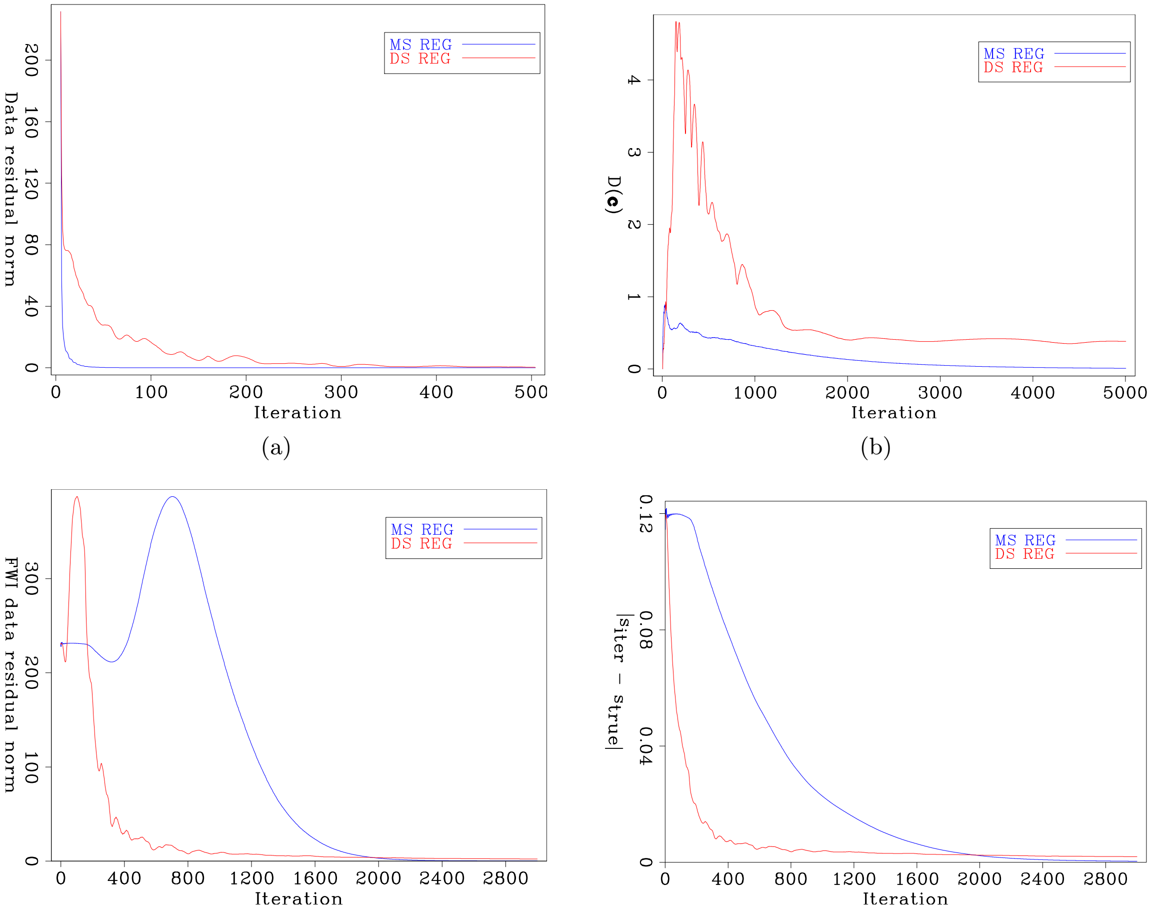
Another approach of regularization is to add a data-space regularization term. One way to write this objective is as:

$$J_{\mathrm{D}}(s, \mathbf{c}) = \frac{1}{2}\left\|\tilde{\mathcal{L}}(s, \mathbf{c}) - \mathbf{d}_r\right\|_2^2 + \frac{\epsilon}{2}\left\|\tilde{\mathcal{L}}(s, \mathbf{c}) - \tilde{\mathcal{L}}(s, \mathbf{F}\mathbf{c})\right\|_2^2. \tag{20}$$

As was discussed with the model-space regularization, choosing $\mathbf{F}$ to be the DSO operator provides the best convergence properties for optimizing this objective function. The results of running this inversion for 10000 iterations and a starting model of $s_0 = 1.12$ s/km and $\mathbf{c}_0 = \mathbf{0}$ with $\epsilon = 10$ and the DSO operator as the focusing operator $\mathbf{F}$ are shown in Figure 5.

Comparison

Comparing both model and data space regularization results (blue and red curves respectively) in each of the panels in Figure 5 we observe from the data residual norms (panel (a)) that the model space regularization reduces the extended FWI objective function ($||\tilde{\mathcal{L}}(s, \mathbf{c}) - \mathbf{d}_r||_2^2$) faster than does the data space. This is due to the fact that with the model space regularization, we are directly focusing the model therefore providing better focusing of the extended model ($\mathbf{c}$). This is clear in panel (b) which shows the focusing measure $D(\mathbf{c})$ with iteration. However, the data space regularization does update the model much faster than the model space regularization. This is evident in panels (c) and (d) where the FWI data residual norm ($\frac{1}{2}||\mathcal{L}(s) - \mathbf{d}_r||_2^2$) and the model residual are shown respectively. From



(a) (b)

(c) (d)

Figure 5: Results of inversions with model space (blue curve) and data space (red curve) regularization. Both inversions ran for 7000 iterations and plots are windowed here for display. The inversions were run with starting a model of $s_0 = 1.12$ s/km and $\mathbf{c}_0 = 0$. (a) Data residual norms of extended FWI objective function ($\frac{1}{2}||\tilde{\mathcal{L}}\left(s, \mathbf{c}\right) - \mathbf{d}_r||_2^2$), (b) focusing measure $D(\mathbf{c})$ as defined in equation 7, (c) data residual norms of FWI objective function ($\frac{1}{2}||\mathcal{L}\left(s\right) - \mathbf{d}_r||_2^2$), and (d), model residual ($|s_{\text{iter}} - \bar{s}|$ where s_{iter} is the physical slowness at the current iteration). [**CR**] biondo1/joefig RN-comp,FM-comp,FWIR-comp,SMN-comp

these figures it is clear that in only 600 iterations the updated physical slowness is within 0.45% of the true slowness. In contrast, the model space regularization needs more than 1500 iterations before it reaches that point in the inversion. At later iterations, it is evident in both panels (c) and (d) that the model-space regularization does reduce the objective function more than the data-space regularization. This only occurs when the residuals are quite small.

These results suggest that a data-space regularization with frequent restarts; that is, by resetting set **c** to zero might be the best approach within this class of methods. Biondi and Almomin (2014) presented a nested optimization algorithm that included a restart at each outer iteration. At the limit, if the algorithm is restarted by resetting set **c** to zero after each update of the slowness model, we will implement an inversion scheme close the the alternating algorithms presented in the next section.

A CLASS OF ALTERNATING ALGORITHMS

In this section, we discuss a class of algorithms to solve the extended FWI problem. We begin by first introducing an algorithm based on a simple intuitive idea, which we will refer to as the "alternating algorithm". Guided by computational evidence that the algorithm converges to the true slowness $\bar{s}$ for any incorrect starting slowness s_0, we attempt to understand the underlying cause of its global convergence. Mathematical analysis reveals the key property governing its convergence properties. We discuss some of these findings briefly.

Although we are not able to mathematically prove global convergence of the alternating algorithm, we are able to build on the analysis to develop a simpler modified algorithm that retains the global convergence properties. We will refer to it as the "modified alternating algorithm". It should be noted that the modified alternating algorithm is under investigation and it converges faster, but this property may be specific to the 1D problem analyzed in this paper. Whether these properties hold for the general case is not known at this time.

Throughout this section we do not enforce the requirement that $c_0 = 0$. It is not necessary for the class of algorithms discussed in this section for the specified 1D problem. In addition, we also assume that the length of the convolution filter is given by $N_\tau = 2N_t - 1$.

Alternating algorithm

We motivate the first algorithm by considering the residual of the total modeled data using the extended modeling operator measured in the l^2 norm. For a given starting slowness s_0 and convolution filter **c**, we denote this quantity as $J_{\mathbf{c}}(s_0, \mathbf{c})$. It is defined below as:

$$J_{\mathbf{c}}(s_0, \mathbf{c}) = \frac{1}{2} || \widetilde{\mathcal{L}}(s_0, \mathbf{c}) - \mathbf{d}_r ||_2^2. \tag{21}$$

It is important to remember that we had originally introduced the convolution filter **c** to represent an extended set of model parameters that we could change as we like to model the recorded data, for any starting slowness s_0. A natural way to achieve this goal is to fix s_0 in equation 21, and then perform gradient descent to determine a suitable **c** that minimizes $J_{\mathbf{c}}(s_0, \mathbf{c})$. In fact for fixed s_0, the function $J_{\mathbf{c}}(s_0, \mathbf{c})$ is a semidefinite quadratic, and thus doing a sequence of steepest descent iterations over $\mathbf{c}$ will converge to a global minimum, which in absence of noise in the data happens to be zero. However, solving such a subproblem to convergence involves repeated iterations, irrespective of the optimization algorithm employed.

We look at a cheaper alternative which is to only look at the negative gradient at the first iteration, still for fixed s_0 but with $\mathbf{c} = \mathbf{0}$. We denote this quantity $\hat{\mathbf{c}}(s_0)$ and define it below,

$$\hat{\mathbf{c}}(s_0) = - \left. \frac{\partial J_{\mathbf{c}}(s, \mathbf{c})}{\partial \mathbf{c}} \right|_{s=s_0, \mathbf{c}=\mathbf{0}} = \tilde{\mathbf{L}}'(s_o) \left[\mathbf{d}_r - \widetilde{\mathcal{L}}(s_0, \mathbf{0}) \right] = \tilde{\mathbf{L}}'(s_o) \left[\mathbf{d}_r - \mathcal{L}(s_0) \right]. \tag{22}$$

Note that $-\hat{\mathbf{c}}(s_0)$ is the same quantity that was plotted before in Figure 1d, i.e the application of $\tilde{\mathbf{L}}'(s_o)$ to the data residual $\mathcal{L}(s_0) - \mathbf{d}_r$. This figure can be understood as a superposition of a central positive band that is invariant with respect to s_0, and a diagonal negative band that depends on s_0. The positive band is exactly the term $\tilde{\mathbf{L}}'(s_o) \mathcal{L}(s_0)$, while the negative band is the term $-\tilde{\mathbf{L}}'(s_o) \mathbf{d}_r$ appearing in the expression for $\hat{\mathbf{c}}(s_0)$ in equation 22.

It can be mathematically proved that the positive band is invariant with respect to s_0. It is also clear from the figure that the closer s_0 is to $\bar{s}$, the diagonal negative band is closer to the zero lag coefficient of $\hat{\mathbf{c}}(s_0)$. This observation is key in understanding this algorithm and motivates the following idea : given any $\hat{\mathbf{c}}(s_0)$, we can try to apply a focusing operator $\mathbf{F}$ to transform $\hat{\mathbf{c}}(s_0)$ to an approximation of $\hat{\mathbf{c}}(s_0 + \Delta s)$, where Δs is a slowness perturbation towards $\bar{s}$. Mathematically, this idea is expressed below:

$$\mathbf{F}\hat{\mathbf{c}}(s_0) \approx \hat{\mathbf{c}}(s_0 + \Delta s). \tag{23}$$

Assuming that we have carried out the above transformation, the only thing remaining to do is to find a way to recover Δs from $\mathbf{F}\hat{\mathbf{c}}(s_0)$. An intuitive idea would be to match the quantities $\widetilde{\mathcal{L}}(s_0, \hat{\mathbf{c}}(s_0))$ and $\widetilde{\mathcal{L}}(s, \mathbf{F}\hat{\mathbf{c}}(s_0))$ in the least squares sense, where s is close to s_0. For fixed s_0, this leads to the following objective function:

$$J_s(s) = \frac{1}{2} || \widetilde{\mathcal{L}}(s_0, \hat{\mathbf{c}}(s_0)) - \widetilde{\mathcal{L}}(s, \mathbf{F}\hat{\mathbf{c}}(s_0)) ||_2^2. \tag{24}$$

The quantity $\widetilde{\mathcal{L}}(s_0, \hat{\mathbf{c}}(s_0))$ represents the total modeled data using the extended modeling operator for slowness s_0 and convolution filter $\hat{\mathbf{c}}(s_0)$, while $\widetilde{\mathcal{L}}(s, \mathbf{F}\hat{\mathbf{c}}(s_0))$ represents the total modeled data for any slowness value s close to s_0 and the focused convolution filter $\mathbf{F}\hat{\mathbf{c}}(s_0)$. The expectation is that minimizing the objective function $J_s(s)$ will yield a slowness perturbation Δs towards $\bar{s}$. Putting all these ideas together, we have the following algorithm:

1. For any starting slowness s_0, compute $\hat{\mathbf{c}}(s_0) = \tilde{\mathbf{L}}'(s_o) \left[\mathbf{d}_r - \mathcal{L}(s_0) \right].$

2. Start from $s = s_0$, and solve the following local optimization problem:

$$s_* = \underset{s}{\operatorname{argmin}} \;\; J_s(s) \tag{25}$$

3. Set $s_0 = s_*$ and iterate 1-3 till convergence.

In the above discussion $\mathbf{F}$ is a general focusing operator. However in this section, we restrict ourselves to the particular case of the shift focusing operator, which was introduced earlier in equation 18. Note that the choice of the shift focusing operator is inherently restrictive, because we only shift the coefficients in the extended model vector by one sample towards the zero lag coefficient. Doing that automatically enforces small changes in Δs. It is indeed possible to incorporate bigger shifts up to a limit into $\mathbf{F}_S$, and we would obtain similar results to what we present next. The choice of studying the one sample shift focusing operator captures all the effects that would be true with bigger shifts. The subsequent analysis of the alternating algorithm also holds in this regime in a slightly modified form. In fact, with bigger shifts the rate of convergence is much faster at each iteration when s_0 is far away from $\bar{s}$. However, if the shifts are too large we lose the property that Δs is a slowness perturbation towards $\bar{s}$ for all starting slowness s_0.

Such ideas of incorporating bigger shifts and speeding up convergence can also be incorporated with the use of other types of focusing operators like the shrink focusing operator $\mathbf{F}_\alpha$ introduced earlier.

Numerical results

We provide computational evidence that the alternating algorithm converges to $\bar{s}$ for any starting slowness s_0. To illustrate this we start by plotting the objective function $J_s(s)$ for different values of s_0 in Figure 6. It is clearly seen from each panel that minimizing $J_s(s)$ starting from s_0 will yield a slowness update Δs towards $\bar{s}$. Therefore, if this process is repeated at every iteration, we will reach $\bar{s}$. The iterates eventually stop changing when $s_0 = \bar{s}$ as the gradient of $J_s(s)$ with respect to s at $s = \bar{s}$ becomes zero. This test shows that the alternating algorithm converges to the true slowness $\bar{s}$ starting from any s_0.

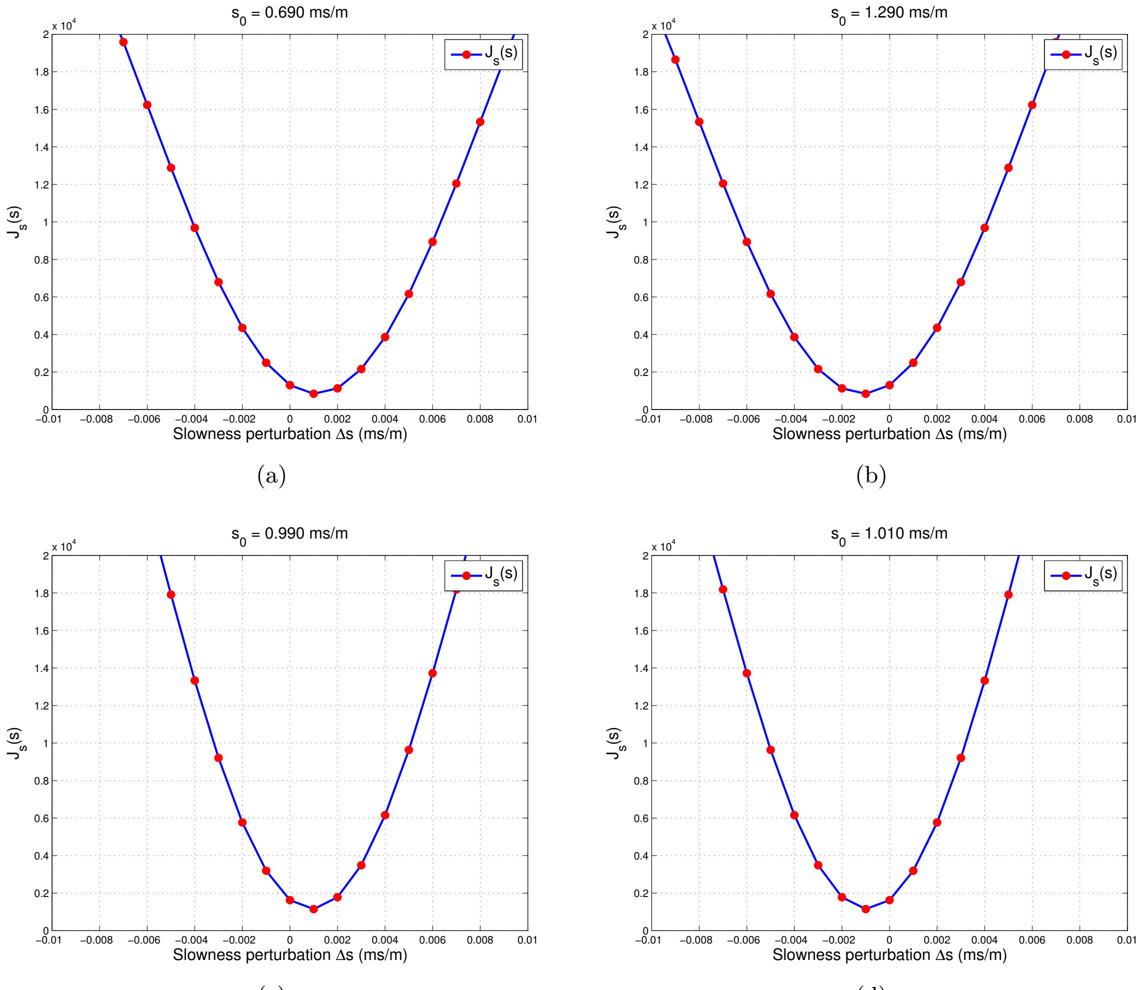
Another way of seeing the global convergence property is to evaluate the gradient of the objective function $J_s(s)$ for each s_0 at $\Delta s = 0$, i.e $\frac{\partial J_s(s_0)}{\partial s}$. We first calculate the gradient of $J_s(s)$ below:

$$\begin{aligned}
\frac{\partial J_s(s)}{\partial s} &= -\left[\frac{\partial \mathcal{L}(s)}{\partial s} + \frac{\partial \tilde{\mathbf{L}}\,(s)}{\partial s}\mathbf{F}\hat{\mathbf{c}}(s_0))\right]' \left[\widetilde{\mathcal{L}}(s_0, \hat{\mathbf{c}}(s_0)) - \widetilde{\mathcal{L}}(s, \mathbf{F}\hat{\mathbf{c}}(s_0))\right] \\
&= -\left[\mathbf{L}(s) + \dot{\tilde{\mathbf{L}}}\,(s)\,\mathbf{F}\hat{\mathbf{c}}(s_0))\right]' \left[\widetilde{\mathcal{L}}(s_0, \hat{\mathbf{c}}(s_0)) - \widetilde{\mathcal{L}}(s, \mathbf{F}\hat{\mathbf{c}}(s_0))\right] \\
&= -\left[\mathbf{L}'(s) + \hat{\mathbf{c}}'(s_0)\mathbf{F}'\,\dot{\tilde{\mathbf{L}}}'\,(s)\right] \left[\widetilde{\mathcal{L}}(s_0, \hat{\mathbf{c}}(s_0)) - \widetilde{\mathcal{L}}(s, \mathbf{F}\hat{\mathbf{c}}(s_0))\right] \quad .
\end{aligned} \tag{26}$$

Evaluating this quantity at $s = s_0$ gives:

$$\frac{\partial J_s(s_0)}{\partial s} = -\left[\mathbf{L}'(s_0) + \hat{\mathbf{c}}'(s_0)\mathbf{F}'\,\dot{\tilde{\mathbf{L}}}'\,(s_o)\right]\tilde{\mathbf{L}}\,(s_o)\,[\mathbf{I}-\mathbf{F}]\hat{\mathbf{c}}(s_0) \quad . \tag{27}$$

We have plotted the quantity $\frac{\partial J_s(s_0)}{\partial s}$ as a function of different starting slowness values s_0 in Figure 7. The figure tells us that for $s_0 > \bar{s}$, the gradient is always positive, while for $s_0 < \bar{s}$, the gradient is always negative. Thus, the search direction which is the negative of the gradient always points in the correct direction. Finally when $s_0 = \bar{s}$, the gradient is zero, which means that the algorithm will terminate when $s_0 = \bar{s}$.



(a) (b)

(c) (d)

Figure 6: Plot of the function $J_s(s)$ for different values of s_0 : (a) $s_0 = 0.69$ ms/m represents the case when s_0 is slow compared to $\bar{s}$, (b) $s_0 = 1.29$ ms/m represents the case when s_0 is fast compared to $\bar{s}$, (c) $s_0 = 0.99$ ms/m represents the case when s_0 is almost close to $\bar{s}$ on the slower side, and (d) $s_0 = 1.01$ ms/m represents the case when s_0 is almost close to $\bar{s}$ on the faster side. [**ER**] biondo1/rahulfig Js-50,Js-110,Js-80,Js-82

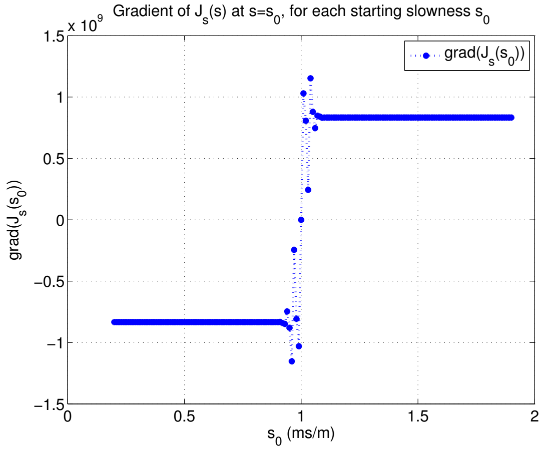


Figure 7: Plot of the quantity $\frac{\partial J_s(s_0)}{\partial s}$ in equation 27 as a function of different starting slowness values s_0. [**ER**] biondo1/rahulfig Jsgrad

Analysis of the alternating algorithm

The computational evidence of global convergence to $\bar{s}$ naturally leads us to take a closer look at the alternating algorithm. Here the main question of interest is *"why is the algorithm converging ?"*. We want to isolate the pieces of the objective function that is responsible for its convergence.

We first focus our attention to the behavior of $J_s(s)$ for some fixed s_0. We will also henceforth use $\hat{\mathbf{c}}$ to denote $\hat{\mathbf{c}}(s_0)$, unless stated otherwise to avoid excessive notation. We start by expanding the expression for $J_s(s)$ and group similar terms to get the following:

$$\begin{aligned} J_s(s) &= \frac{1}{2}||\widetilde{\mathcal{L}}(s_0,\hat{\mathbf{c}}) - \widetilde{\mathcal{L}}(s,\mathbf{F}\hat{\mathbf{c}})||_2^2 = \frac{1}{2}||\left[\mathcal{L}(s_0)-\mathcal{L}(s)\right] + \left[\tilde{\mathbf{L}}\left(s_o\right)\hat{\mathbf{c}} - \tilde{\mathbf{L}}\left(s\right)\mathbf{F}\hat{\mathbf{c}}\right]||_2^2 \\ &= \frac{1}{2}||\left[\mathcal{L}(s_0)-\mathcal{L}(s)\right] - \left[\tilde{\mathbf{L}}\left(s_o\right)\tilde{\mathbf{L}}'\left(s_o\right)\mathcal{L}(s_0) - \tilde{\mathbf{L}}\left(s\right)\mathbf{F}\tilde{\mathbf{L}}'\left(s_o\right)\mathcal{L}(s_0)\right] \\ &+ \left[\tilde{\mathbf{L}}\left(s_o\right)\tilde{\mathbf{L}}'\left(s_o\right)\mathbf{d}_r - \tilde{\mathbf{L}}\left(s\right)\mathbf{F}\tilde{\mathbf{L}}'\left(s_o\right)\mathbf{d}_r\right]||_2^2 \\ &= \frac{1}{2}||\mathbf{u}+\mathbf{v}+\mathbf{w}||_2^2 = \frac{1}{2}\left[||\mathbf{u}||_2^2 + ||\mathbf{v}||_2^2 + ||\mathbf{w}||_2^2\right] + \left[\mathbf{u}'\mathbf{v} + \mathbf{v}'\mathbf{w} + \mathbf{w}'\mathbf{u}\right] \quad , \end{aligned} \tag{28}$$

where we have denoted,

$$\begin{aligned} \mathbf{u} &:= \mathbf{u}(s) = \mathcal{L}(s_0) - \mathcal{L}(s) \\ \mathbf{v} &:= \mathbf{v}(s) = -[\tilde{\mathbf{L}}\left(s_o\right) - \tilde{\mathbf{L}}\left(s\right)\mathbf{F}]\tilde{\mathbf{L}}'\left(s_o\right)\mathcal{L}(s_0) \\ \mathbf{w} &:= \mathbf{w}(s) = [\tilde{\mathbf{L}}\left(s_o\right) - \tilde{\mathbf{L}}\left(s\right)\mathbf{F}]\tilde{\mathbf{L}}'\left(s_o\right)\mathbf{d}_r \quad . \end{aligned} \tag{29}$$

The different terms involving u,v and w

It turns out that the most interesting term is the one involving only $||\mathbf{w}||_2^2$. This term is the key in getting the correct update when s_0 is far away from $\bar{s}$, and so the modeled data

$\mathcal{L}(s_0)$ and the recorded data $\mathbf{d}_r$ don't interfere with each other. When this is the case, two key mathematical properties hold for the 1D problem being studied, as stated below:

$$\mathbf{F}\tilde{\mathbf{L}}'(s_o)\,\mathbf{d}_r = \tilde{\mathbf{L}}'(s_0+\Delta s)\,\mathbf{d}_r + \mathcal{O}((\delta t)^2) \quad , \tag{30a}$$

$$\tilde{\mathbf{L}}(s_o)\,\tilde{\mathbf{L}}'(s_o) = \tilde{\mathbf{L}}(s_0+\Delta s)\,\tilde{\mathbf{L}}'(s_0+\Delta s) + \mathcal{O}((\delta t)^2) \quad . \tag{30b}$$

Demonstration of equations 30 is straightforward but lengthy; therefore, we decided to omit it from the text.

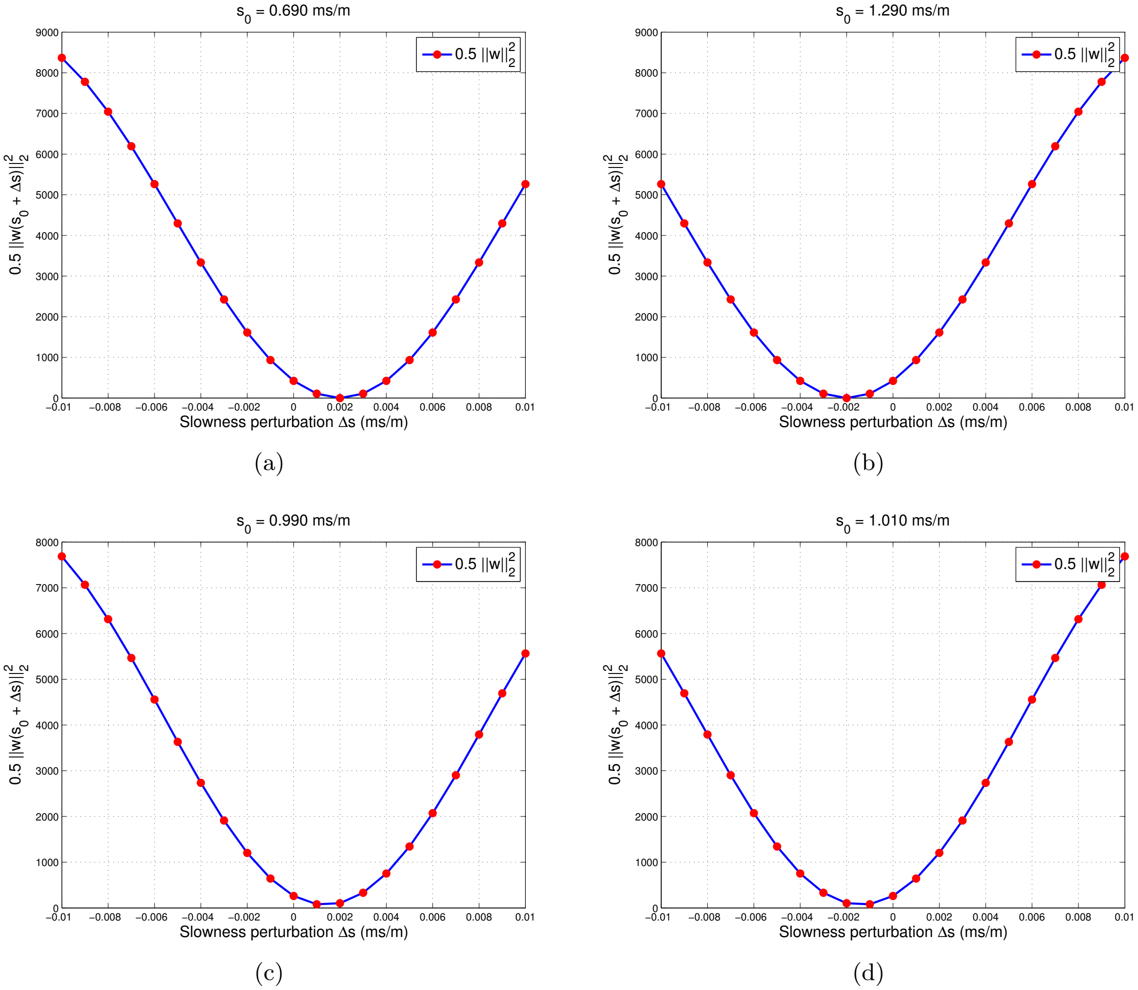
In equation 30, Δs is a slowness perturbation in the correct direction (towards $\bar{s}$). The immediate consequence of these relations is that the quantity $\frac{1}{2}||\mathbf{w}||_2^2$ vanishes at $s_0+\Delta s$. This is illustrated below in equation 31, where the second line follows from 30(a) and the third line follows from 30(b).

$$\begin{aligned}
\frac{1}{2}||\mathbf{w}||_2^2\Big|_{s_0+\Delta s} &= \frac{1}{2}||\tilde{\mathbf{L}}(s_0+\Delta s)\,\mathbf{F}\tilde{\mathbf{L}}'(s_o)\,\mathbf{d}_r - \tilde{\mathbf{L}}(s_o)\,\tilde{\mathbf{L}}'(s_o)\,\mathbf{d}_r||_2^2 \\
&= \frac{1}{2}||\tilde{\mathbf{L}}(s_0+\Delta s)\,\tilde{\mathbf{L}}'(s_0+\Delta s)\,\mathbf{d_r} - \tilde{\mathbf{L}}(s_o)\,\tilde{\mathbf{L}}'(s_o)\,\mathbf{d}_r + \mathcal{O}((\delta t)^2)||_2^2 \\
&= \frac{1}{2}||\tilde{\mathbf{L}}(s_o)\,\tilde{\mathbf{L}}'(s_o)\,\mathbf{d}_r - \tilde{\mathbf{L}}(s_o)\,\tilde{\mathbf{L}}'(s_o)\,\mathbf{d}_r + \mathcal{O}((\delta t)^2)||_2^2 \\
&= \mathcal{O}((\delta t)^4) \approx 0 \quad .
\end{aligned} \tag{31}$$

Numerical studies also suggest that the function $\frac{1}{2}||\mathbf{w}||_2^2$ is increasing in the interval $[\min(s_0, s_0+\Delta s), \max(s_0, s_0+\Delta s)]$. Thus if we carry out local minimization of the $\frac{1}{2}||\mathbf{w}||_2^2$ term starting from the initial slowness s_0, we would obtain the new optimal point $s_0+\Delta s$, which is always closer to the true slowness $\bar{s}$. It is to be noted that the above argument only holds in the regime when s_0 is far away from $\bar{s}$. When this is not the case, the modeled data $\mathcal{L}(s_0)$ and the recorded data $\mathbf{d}_r$ begin to interfere and equation 31 does not hold. However, numerical studies seem to indicate that one can still get the correct update direction by minimizing $\frac{1}{2}||\mathbf{w}||_2^2$.

These aspects are illustrated in Figure 8, where we have plotted the quantity $\frac{1}{2}||\mathbf{w}(s_0+\Delta s)||_2^2$ as a function of the slowness perturbation Δs around s_0, for different values of s_0. Figures 8(a) and 8(b) represent cases when s_0 and $\bar{s}$ are sufficiently far apart so that there is no interference between $\mathcal{L}(s_0)$ and $\mathbf{d}_r$. As can be seen on both the figures, the function goes to zero on the "correct" side, i.e if we start from $\Delta s = 0$ and try to minimize $\frac{1}{2}||\mathbf{w}(s_0+\Delta s)||_2^2$, the optimal Δs would represent a step towards $\bar{s}$. The same fact is true also for Figures 8(c) and 8(d), which represent cases where s_0 is so close to $\bar{s}$ that there is interference between $\mathcal{L}(s_0)$ and $\mathbf{d}_r$. In both of these cases, one can see that the minima of $\frac{1}{2}||\mathbf{w}(s_0+\Delta s)||_2^2$ is still in the right direction, but the function does not become zero at the minima. This is precisely connected to the fact that equation 31 is losing accuracy in this regime.

It seems from the analysis of the $\frac{1}{2}||\mathbf{w}||_2^2$ term that if one ignored all the other terms in the expression for $J_s(s)$, we would still have global convergence. In fact this is actually the



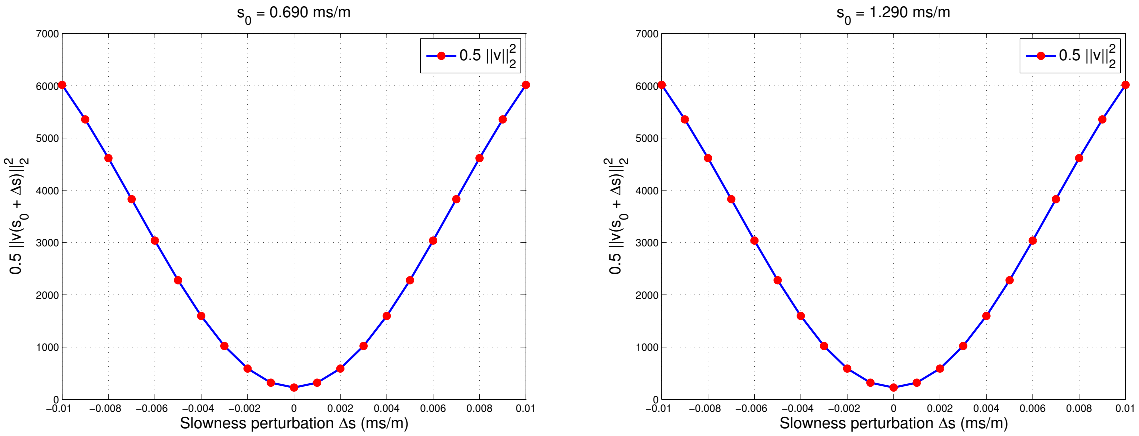
(a) (b) (c) (d)

Figure 8: Plot of the function $\frac{1}{2}||\mathbf{w}(s_0 + \Delta s)||_2^2$ for different values of s_0 : (a) $s_0 = 0.69$ ms/m represents the case when s_0 is slow compared to $\bar{s}$, (b) $s_0 = 1.29$ ms/m represents the case when s_0 is fast compared to $\bar{s}$, (c) $s_0 = 0.99$ ms/m represents the case when s_0 is almost close to $\bar{s}$ on the slower side, and (d) $s_0 = 1.01$ ms/m represents the case when s_0 is almost close to $\bar{s}$ on the faster side. [**ER**] biondo1/rahulfig W-50,W-110,W-80,W-82

case, and would later motivate the development of the "modified alternating algorithm". However, we still need to consider the behavior of the remaining terms in the expression for $J_s(s)$ and understand if they help or impede the convergence. We thus turn to the $\frac{1}{2}||\mathbf{u}||_2^2$ and $\frac{1}{2}||\mathbf{v}||_2^2$ terms next.

The $\frac{1}{2}||\mathbf{u}||_2^2$ term is not useful. Using the definition of $\mathbf{u}$ in equation 29, we have $\frac{1}{2}||\mathbf{u}||_2^2 = \frac{1}{2}||\mathcal{L}(s) - \mathcal{L}(s_0)||_2^2$. This looks exactly like the FWI objective function $\frac{1}{2}||\mathcal{L}(s) - \mathbf{d}_r||_2^2$, but where the recorded data $\mathbf{d}_r = \mathcal{L}(\bar{s})$ has been replaced by $\mathcal{L}(s_0)$, i.e data modeled using an incorrect slowness s_0. We know from the properties of the FWI objective function that it is a convex function in a sufficiently close neighborhood of the true slowness. Therefore, this property also carries over to the $\frac{1}{2}||\mathbf{u}||_2^2$ term, and we conclude that the $\frac{1}{2}||\mathbf{u}||_2^2$ is locally convex around s_0. This means that if we start sufficiently close to s_0, the gradient of s with respect to the $\frac{1}{2}||\mathbf{u}||_2^2$ term will always point towards s_0. This observation is important, as it says that while the presence of the $\frac{1}{2}||\mathbf{w}||_2^2$ term in $J_s(s)$ will provide the correct update direction, including the $\frac{1}{2}||\mathbf{u}||_2^2$ term in $J_s(s)$ will oppose the correct slowness update. Clearly, this is not a desirable property and we thus conclude that we should omit the $\frac{1}{2}||\mathbf{u}||_2^2$ term in the definition of $J_s(s)$.

Finally we consider the $\frac{1}{2}||\mathbf{v}||_2^2$ term. It turns out that this term is also convex locally around s_0. For this particular 1D case, an interesting fact about this term is that its profile does not vary with s_0. This is illustrated in Figure 9. As clearly seen, the curves in both the figures are exactly same. In fact, this is also true for all possible values of s_0 irrespective of whether s_0 is close or far from $\bar{s}$, but this is most likely only true for the specific 1D problem we are studying. However, we think that the local convexity property of $\frac{1}{2}||\mathbf{v}||_2^2$ close to s_0 may be more general. Thus, presence of this term will also impede convergence as its gradient will oppose any change from s_0.



(a) (b)

Figure 9: Plot of the function $\frac{1}{2}||\mathbf{v}(s_0+\Delta s)||_2^2$ for different values of s_0 : (a) $s_0 = 0.69$ ms/m corresponds to the case when s_0 is slower compared to $\bar{s}$, and (b) $s_0 = 1.29$ ms/m corresponds to the case when s_0 is faster compared to $\bar{s}$. [**ER**] biondo1/rahulfig V-50,V-110

The cross terms $\mathbf{u}'\mathbf{v}+\mathbf{v}'\mathbf{w}+\mathbf{w}'\mathbf{u}$ are quite complicated and evades mathematical analysis at the moment. Preliminary analysis has not revealed any clear understanding of how they may be affecting convergence. In general, they produce gradients which are sometimes in the correct direction and at other times not. So it is quite a miracle that the full expression for $J_s(s)$ produces the correct update direction, as evidenced in the numerical results shown earlier.

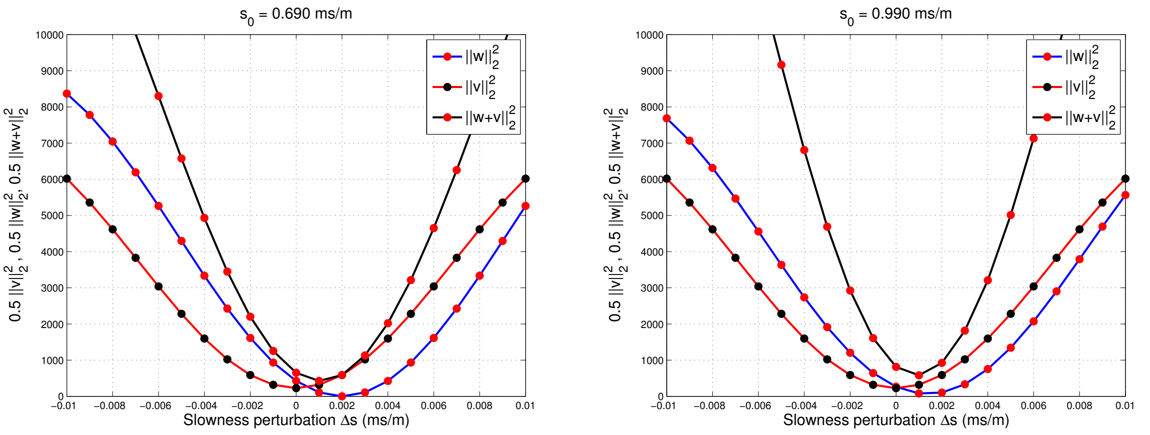
Slowdown of convergence

It is quite clear that for this 1D problem, the effect of the $\frac{1}{2}||\mathbf{w}||_2^2$ term dominates that of the others in the full expression for $J_s(s)$. So as a net, we obtain the correct slowness perturbation Δs by minimizing $J_s(s)$ at every step of the iterative algorithm. Starting from any slowness s_0, one step of the algorithm takes us closer to the true slowness $\bar{s}$, and this process iterated enough times finally takes us to $\bar{s}$. The process converges to $\bar{s}$ because as we get closer and closer to $\bar{s}$, the updates Δs get smaller and smaller. In fact for $s_0 = \bar{s}$, we have $\frac{\partial J_s(\bar{s})}{\partial s} = 0$, and so we do not get any more update.

However from the point of view of algorithmic efficiency and rate of convergence, we would really like to avoid the ill-effects associated with the terms involving $\mathbf{u}$ and $\mathbf{v}$. Based on the analysis done so far, we would like to only use the $\frac{1}{2}||\mathbf{w}||_2^2$ term for solving the local minimization problem at each iteration. But first, we provide some numerical evidence of how the convergence is affected. As an example, we have chosen to drop all terms involving $\mathbf{u}$ from the expression of $J_s(s)$, and analyze the behavior of the remaining terms given by $J_s(s) = \frac{1}{2}||\mathbf{w}||_2^2 + \frac{1}{2}||\mathbf{v}||_2^2 + \mathbf{v}'\mathbf{w} = \frac{1}{2}||\mathbf{w}+\mathbf{v}||_2^2$. This situation is plotted in Figures 10(a),10(b) for slowness values $s_0 = 0.69$ ms/m, 0.99 ms/m. In each figure, we have displayed the quantities $\frac{1}{2}||\mathbf{v}||_2^2$, $\frac{1}{2}||\mathbf{w}||_2^2$ and $\frac{1}{2}||\mathbf{w}+\mathbf{v}||_2^2$. As we can clearly see in both cases, minimizing $\frac{1}{2}||\mathbf{w}+\mathbf{v}||_2^2$ or $\frac{1}{2}||\mathbf{w}||_2^2$ starting from s_0 will yield a slowness perturbation towards $\bar{s}$. But in the latter case, the step will be larger compared to the former. It should be mentioned that similar conclusions are also obtained for the cases when s_0 is faster than $\bar{s}$. We thus conclude that minimizing the complete expression for $J_s(s)$ will take longer to converge to $\bar{s}$ in the presence of the $\mathbf{u}$ and $\mathbf{v}$ terms.

Modified alternating algorithm

To address the convergence issue, we now discuss a modified alternating algorithm that is based on the idea of replacing the original expression for $J_s(s)$ in equation 24 by $J_s(s) = \frac{1}{2}||\mathbf{w}||_2^2 = \frac{1}{2}||[\tilde{\mathbf{L}}\,(s_o) - \tilde{\mathbf{L}}\,(s)\,\mathbf{F}]\tilde{\mathbf{L}}'\,(s_o)\,\mathbf{d}_r||_2^2$. As we have shown previously, local minimization of the $\frac{1}{2}||\mathbf{w}||_2^2$ term produces the correct update direction at every iteration. We had also argued previously based on Figures 8(c) and 8(d) that minimizing the $\frac{1}{2}||\mathbf{w}||_2^2$ term gives the correct update direction, even when s_0 is close to $\bar{s}$. This is indeed the case for the choice of the modeling parameters like wavelet and sampling interval used to create those plots. However, it has also been observed that for other choices of modeling parameters, the $\frac{1}{2}||\mathbf{w}||_2^2$ term starts to have the minima exactly at s_0, when s_0 starts to get close to $\bar{s}$ and the algorithm terminates before reaching $\bar{s}$. It is quite difficult to characterize this situation analytically due to the interference between $\mathcal{L}(s_0)$ and $\mathbf{d}_r$, but we suspect that the effect is related to the shape and frequency of the wavelet and also numerical inaccuracies stemming from the choice of the modeling parameters. We discuss a robust alternative below that was found to not suffer from this issue by modifying the objective function in such a manner that when s_0 is far from $\bar{s}$, it is exactly equal to $\frac{1}{2}||\mathbf{w}||_2^2$, and the only differences are when s_0 is close to $\bar{s}$.



(a) (b)

Figure 10: Plot of the different terms $\frac{1}{2}||\mathbf{v}(s_0+\Delta s)||_2^2$, $\frac{1}{2}||\mathbf{w}(s_0+\Delta s)||_2^2$ and $\frac{1}{2}||\mathbf{w}(s_0+\Delta s)+\mathbf{v}(s_0+\Delta s)||_2^2$ for different values of s_0, indicated by the red, blue and black curves respectively : (a) $s_0 = 0.69$ ms/m represents the case when s_0 is slow compared to $\bar{s}$, and (b) $s_0 = 0.99$ ms/m represents the case when s_0 is almost equal $\bar{s}$, but slightly slower. [**ER**] biondo1/rahulfig VW-50,VW-80

We first define the unit shift operators in the positive and negative directions $\mathbf{S}_+$ and $\mathbf{S}_-$. $\mathbf{S}_+$ defines a linear map $\Re^{N_\tau} \longrightarrow \Re^{N_\tau}$, where each sample is shifted down by one sample. $\mathbf{S}_-$ defines a linear map $\Re^{N_\tau} \longrightarrow \Re^{N_\tau}$, where each sample is shifted up by one sample. These operators have the explicit matrix form as defined below:

$$\mathbf{S}_+ = \begin{bmatrix} 0 & & & & \\ 1 & 0 & & & \\ & \ddots & \ddots & & \\ & & 1 & 0 & \\ & & & 1 & 0 \end{bmatrix}_{N_\tau \times N_\tau} \quad , \quad \mathbf{S}_- = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{bmatrix}_{N_\tau \times N_\tau} \quad . \tag{32}$$

We next introduce the masking operator $\mathbf{M}_i$. It also defines a linear map $\Re^{N_\tau} \longrightarrow \Re^{N_\tau}$, and its action on a vector $\mathbf{c} \in \Re^{N_\tau}$ is defined as follows:

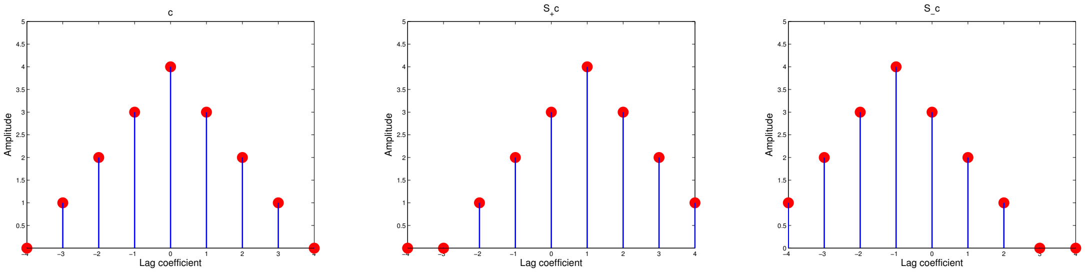
$$\mathbf{M}_i \mathbf{c} = \tilde{\mathbf{c}} \quad \text{, where,} \quad \tilde{c}_j = \begin{cases} c_j, & \text{if } j \leq i \\ 0, & \text{if } j > i \end{cases} \quad . \tag{33}$$

The above quantities $\mathbf{S}_+, \mathbf{S}_-, \mathbf{M}_i$ form the building blocks of the modified algorithm. For the remainder of this section we will redefine the quantity $\hat{\mathbf{c}}$ to be $\hat{\mathbf{c}} = \tilde{\mathbf{L}}'(s_o)\,\mathbf{d}_r$. We also need to define two more quantities $\mathbf{r}_+(s)$ and $\mathbf{r}_-(s)$, which can be interpreted as generalized residuals and defined below:

$$\begin{aligned} \mathbf{r}_+ &:= \mathbf{r}_+(s) = \tilde{\mathbf{L}}(s)\,\mathbf{M}_0 \mathbf{S}_+ \hat{\mathbf{c}} - \tilde{\mathbf{L}}(s_o)\,\mathbf{M}_{-1}\hat{\mathbf{c}} \\ \mathbf{r}_- &:= \mathbf{r}_-(s) = \tilde{\mathbf{L}}(s)\,[\mathbf{I} - \mathbf{M}_{-1}]\mathbf{S}_- \hat{\mathbf{c}} - \tilde{\mathbf{L}}(s_o)\,[\mathbf{I} - \mathbf{M}_0]\hat{\mathbf{c}} \quad . \end{aligned} \tag{34}$$

It is instructive to look at what the combination of terms appearing in equation 34 involving $\mathbf{S}_+, \mathbf{S}_-$ and $\mathbf{M}_0, \mathbf{M}_{-1}$ look like when applied to a vector $\mathbf{c}$. To clarify this point, we choose a dummy vector $\mathbf{c}$ corresponding to the choice of parameters $N_t = 5$, $N_\tau = 2N_t - 1 = 9$ as shown in Figure 11(a). We then applied the unit positive and negative shifts and plotted the results in Figures 11(b) and 11(c), respectively. In Figures 12(a), 12(b), 12(c), and 12(d) we plot the terms $\mathbf{M}_0\mathbf{S}_+\mathbf{c}$, $\mathbf{M}_{-1}\mathbf{c}$, $[\mathbf{I}-\mathbf{M}_{-1}]\mathbf{S}_-\mathbf{c}$ and $[\mathbf{I}-\mathbf{M}_0]\mathbf{c}$, respectively. As seen from these figures, the terms $\mathbf{M}_0\mathbf{S}_+\mathbf{c}$ and $\mathbf{M}_{-1}\mathbf{c}$ are exactly shifted copies of each other with the property that all positive lag coefficients are zero. Similarly, the terms $[\mathbf{I}-\mathbf{M}_{-1}]\mathbf{S}_-\mathbf{c}$ and $[\mathbf{I}-\mathbf{M}_0]\mathbf{c}$ are also shifted copies of each other with the property that all negative lag coefficients are zero.

Thus $\mathbf{r}_+$ term can be interpreted as the difference between the data produced by the linearized extension using the starting slowness s_0 and convolution filter $\mathbf{M}_{-1}\hat{\mathbf{c}}$ (which is the $\tilde{\mathbf{L}}(s_o)\mathbf{M}_{-1}\hat{\mathbf{c}}$ term), and the linearized extension using a slowness s close to s_0 and convolution filter $\mathbf{M}_0\mathbf{S}_+\hat{\mathbf{c}}$ (which is the $\tilde{\mathbf{L}}(s)\mathbf{M}_0\mathbf{S}_+\hat{\mathbf{c}}$ term). A similar interpretation also holds for $\mathbf{r}_-(s)$. It is clear from the definition of the mask operator that when the support of $\hat{\mathbf{c}}$ is strictly negative, $\mathbf{r}_- = \mathbf{0}$ holds identically, and when the support of $\hat{\mathbf{c}}$ is strictly positive, $\mathbf{r}_+ = \mathbf{0}$ holds identically. These situations correspond to the cases $s_0 >> \bar{s}$ and $s_0 << \bar{s}$ respectively and hence the modeled data $\mathcal{L}(s_0)$ and recorded data $\mathbf{d}_r$ do not interfere with each other in both these cases.



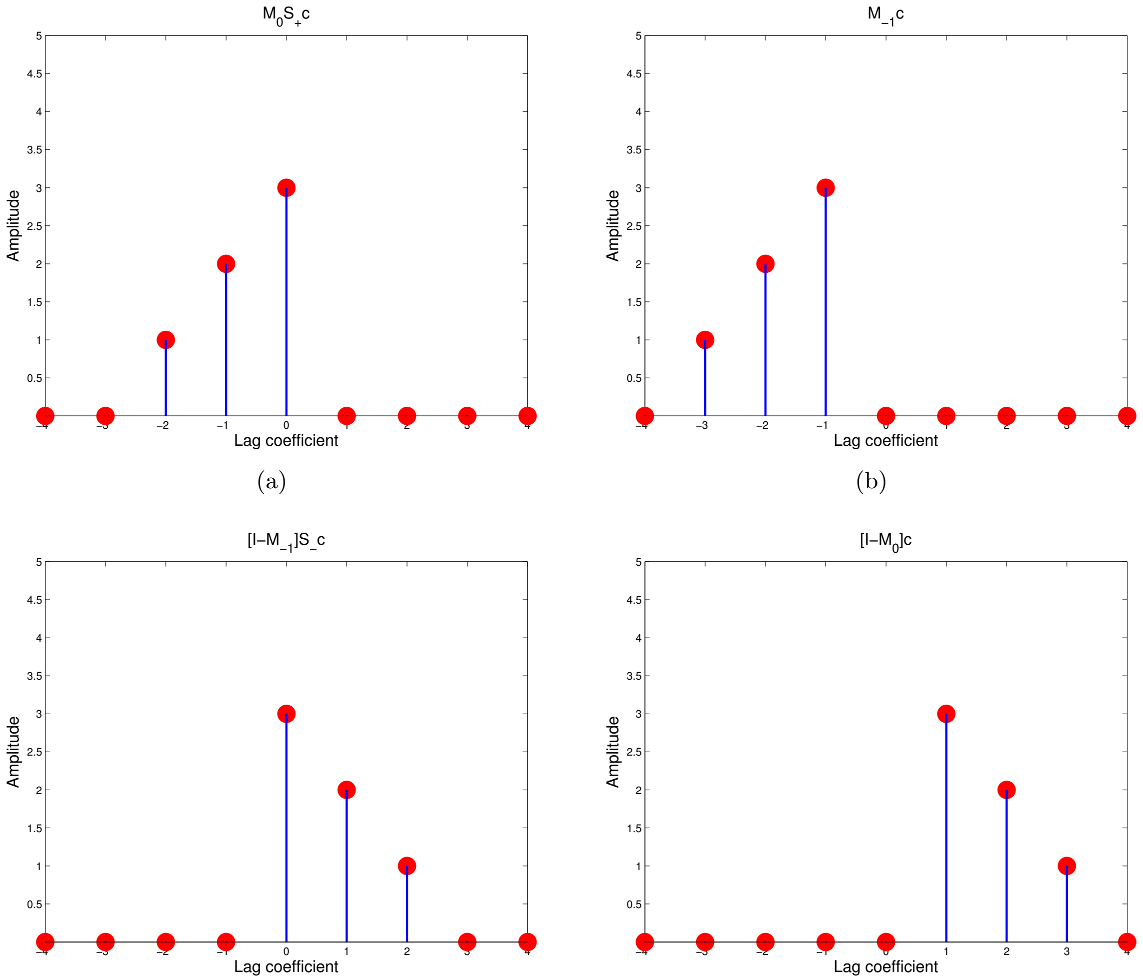
(a) (b) (c)

Figure 11: Plot of the result of application of the shift operators to a vector corresponding to the choice of parameters $N_t = 5$, $N_\tau = 9$. (a) This is the vector $\mathbf{c}$. (b) Result of application of unit positive shift $\mathbf{S}_+\mathbf{c}$. (c) Result of application of unit negative shift $\mathbf{S}_-\mathbf{c}$. [**ER**] biondo1/rahulfig C,SpC,SmC

The preceding observations allow us to define the modified objective function $J_M(s)$ in terms of the squared l^2 norms of $\mathbf{r}_+$ and $\mathbf{r}_-$ which we define to be $J_+(s)$ and $J_-(s)$ respectively, as below:

$$\begin{aligned} J_+(s) &= \frac{1}{2}||\mathbf{r}_+||_2^2 = \frac{1}{2}||\tilde{\mathbf{L}}(s)\,\mathbf{M}_0\mathbf{S}_+\hat{\mathbf{c}} - \tilde{\mathbf{L}}(s_o)\,\mathbf{M}_{-1}\hat{\mathbf{c}}||_2^2 \\ J_-(s) &= \frac{1}{2}||\mathbf{r}_-||_2^2 = \frac{1}{2}||\tilde{\mathbf{L}}(s)\,[\mathbf{I}-\mathbf{M}_{-1}]\mathbf{S}_-\hat{\mathbf{c}} - \tilde{\mathbf{L}}(s_o)\,[\mathbf{I}-\mathbf{M}_0]\hat{\mathbf{c}}||_2^2 \\ J_M(s) &= J_+(s) + J_-(s) \quad . \end{aligned} \tag{35}$$

We can finally write the modified alternating algorithm:



(a) (b) (c) (d)

Figure 12: Plot of the composite terms appearing in equation 34, involving the mask and shift operators, with vector $\mathbf{c}$ in Figure 11(a). (a) $\mathbf{M}_0\mathbf{S}_+\mathbf{c}$ - The result is the zeroing of all samples of $\mathbf{S}_+\mathbf{c}$ after the sample at index 0. (b) $\mathbf{M}_{-1}\mathbf{c}$ - The result is the zeroing of all samples of $\mathbf{c}$ after the sample at index -1. (c) $[\mathbf{I}-\mathbf{M}_{-1}]\mathbf{S}_-\mathbf{c}$ - The result is the zeroing of all samples of $\mathbf{S}_-\mathbf{c}$ before the sample at index 0. (d) $[\mathbf{I}-\mathbf{M}_0]\mathbf{c}$ - The result is the zeroing of all samples of $\mathbf{c}$ before the sample at index 1. [**ER**] biondo1/rahulfig C1,C2,C3,C4

1. $\hat{\mathbf{c}} = \tilde{\mathbf{L}}' \left(s_o\right) \mathbf{d}_r$.
2. Start from $s = s_0$, and solve the local minimization problem:

$$s_* = \underset{s}{\operatorname{argmin}} \;\; J_M(s) \tag{36}$$

3. Set $s_0 = s_*$ and iterate 1-3 till convergence.

The objective function $J_M(s)$ has two interesting properties. The first one is that when s_0 is far away from $\bar{s}$, $J_M(s) = \frac{1}{2}||\mathbf{w}||_2^2$. The second property is that when $s_0 >> \bar{s}$, $J_-(s) = 0$ and hence $J_M(s) = J_+(s)$, and similarly when $s_0 << \bar{s}$, $J_+(s) = 0$ and hence $J_M(s) = J_-(s)$. Both $J_+(s)$ and $J_-(s)$ are non-zero only when $\mathcal{L}(s_0)$ and $\mathbf{d}_r$ are interfering, which happen when s_0 is close to $\bar{s}$. We have plotted the behavior of the functions $J_M(s), J_+(s), J_-(s)$ in Figure 13. It is clear from these plots that the modified alternating algorithm will converge to the true slowness $\bar{s}$ from any starting slowness s_0, as minimizing $J_M(s)$ yields a step Δs towards $\bar{s}$.

We finally note that the modified alternating algorithm converges faster than the alternating algorithm. This can be realized by noting the fact already mentioned that for s_0 sufficiently far away from $\bar{s}$, $J_M(s) = \frac{1}{2}||\mathbf{w}||_2^2$. This observation can be seen by comparing the profiles of $J_M(s)$ in Figures 13(a) and 13(b), with the profiles of $\frac{1}{2}||\mathbf{w}||_2^2$ in Figures 8(a) and 8(b). Thus in the regime when s_0 is far away from $\bar{s}$, the convergence is controlled by the behavior of the $\frac{1}{2}||\mathbf{w}||_2^2$ term. But we know that the $\frac{1}{2}||\mathbf{w}||_2^2$ term has better convergence than $J_s(s)$, and thus so does the modified alternating algorithm.

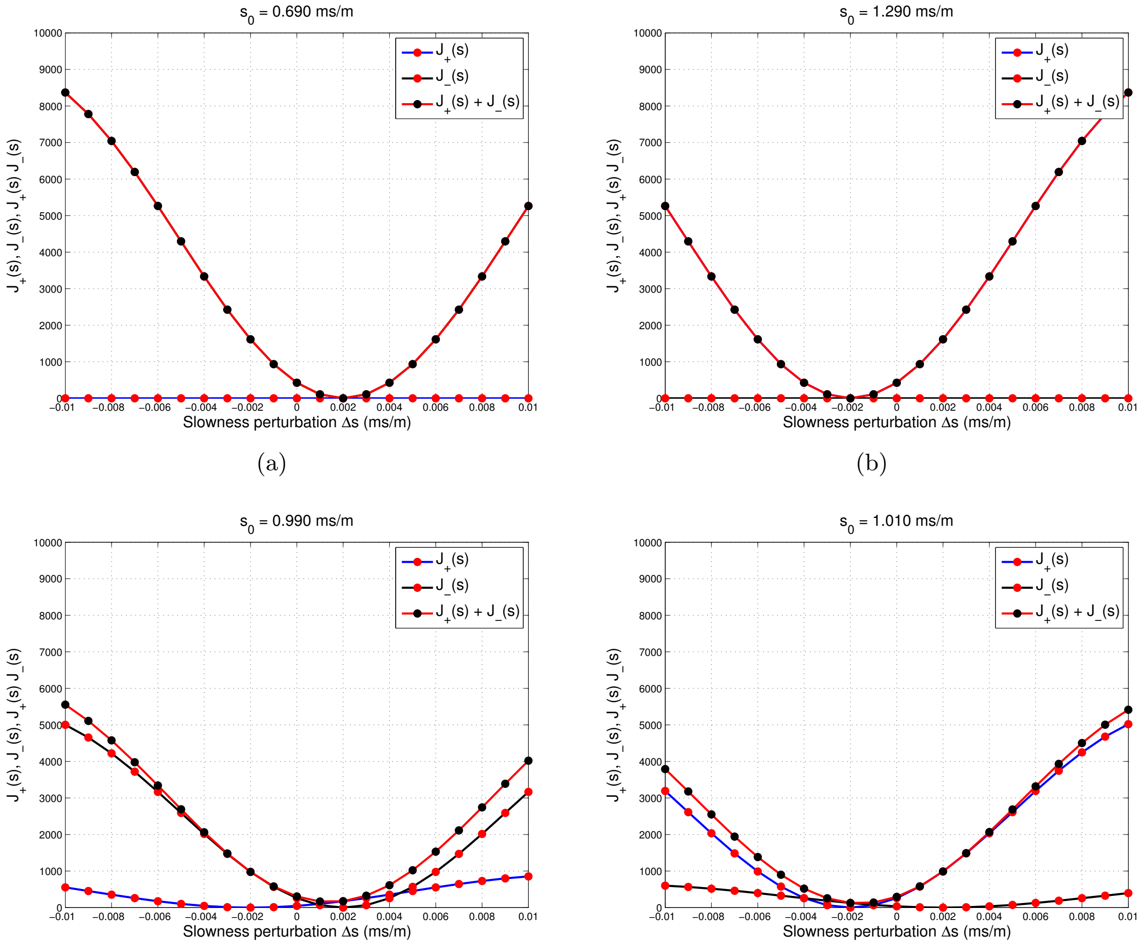
It must be mentioned here that just like it was discussed that it is possible to incorporate the idea of bigger shifts in the shift focusing operator used in the alternating algorithm, it is possible to do the same thing also with the modified alternating algorithm. Doing so will increase the magnitude of the update Δs obtained at each iteration leading to faster convergence, but this will only work up to some maximum shift beyond which the global convergence property will be lost. However, the essential features of running the alternating algorithm with bigger shifts are similar to what have been presented here with the unit shift operators.

FWI-WEMVA OBJECTIVE FUNCTION

Finally we analyze the behavior of the following objective function that incorporates the focusing operator directly in the data fitting term:

$$\begin{aligned} J_{\mathrm{FW}}\left(s\right) &= \frac{1}{2}\left\|\left(\mathbf{I}-\mathbf{F}\right)\tilde{\mathbf{L}}'\left(s\right)\left[\tilde{\mathcal{L}}\left(s,\mathbf{c}=0\right)-\mathbf{d}_r\right]\right\|_2^2 \\ &= \frac{1}{2}\left\|\left(\mathbf{I}-\mathbf{F}\right)\tilde{\mathbf{L}}'\left(s\right)\left[\mathcal{L}\left(s\right)-\mathbf{d}_r\right]\right\|_2^2 . \end{aligned} \tag{37}$$

This objective function is related to the one presented by Symes (2008) in equation 14. One of its attractive properties is that it depends on slowness through both the data residuals $\left(\mathcal{L}\left(s\right)-\mathbf{d}_r\right)$ and the focusing of the backprojection of these residuals into the space of $\mathbf{c}$ by the operator $\tilde{\mathbf{L}}'\left(s\right)$. Our conjecture is that during the inversion process, the gradient component corresponding to the direct dependency on the data residuals introduces short



(a) (b)

(c) (d)

Figure 13: Plot of the functions $J_M(s), J_+(s), J_-(s)$ for different value of starting slowness s_0 : (a) $s_0 = 0.69$ ms/m represents the case when s_0 is slow compared to $\bar{s}$, and hence $J_+(s) = 0$, (b) $s_0 = 1.29$ ms/m represents the case when s_0 is fast compared to $\bar{s}$, and hence $J_-(s) = 0$, (c) $s_0 = 0.99$ ms/m represents the case when s_0 is almost close to $\bar{s}$ on the slower side, and (d) $s_0 = 1.01$ ms/m represents the case when s_0 is almost close to $\bar{s}$ on the faster side. [**ER**] biondo1/rahulfig Jm-50,Jm-110,Jm-80,Jm-82

wavelengths into the slowness model, whereas the gradient component corresponding to the focusing of $\mathbf{c}$ introduces long wavelengths into the slowness model. If that were the case, optimizing this objective function would have the potential of achieving simultaneous inversion of all model scales. Unfortunately, this conjecture cannot be fully tested using our simple 1D model because we assumed the slowness to be a scalar, and obviously cannot be decomposed into different scales.

Another attractive properties of the objective function in equation 37 is that, when the amplitude focusing operators $\mathbf{F}_D$ and $\mathbf{F}_G$ are used, its behavior substantially changes according to the value of the parameter τ_{W}. These changes in behavior of J_{FW} can be easily understood by analyzing the gradient of the objective function with respect to slowness. This gradient has two terms because both $\mathcal{L}$ and $\tilde{\mathbf{L}}'$ are function of the slowness. The term deriving from the dependency of $\mathcal{L}$ from s is a FWI-like gradient, whereas the one deriving from the dependency of $\tilde{\mathbf{L}}'$ from s is a WEMVA-like gradient. The total gradient can be expressed as follows:

$$\begin{aligned}
\nabla J_{\mathrm{FW}} &= \underbrace{\mathbf{L}'\left(s\right)\tilde{\mathbf{L}}\left(s\right)\left(\mathbf{I}-\mathbf{F}\right)'\left(\mathbf{I}-\mathbf{F}\right)\tilde{\mathbf{L}}'\left(s\right)\left[\mathcal{L}\left(s\right)-\mathbf{d}_r\right]}_{\mathrm{FWI-like\ \ gradient}} && (38)\\
&+ \underbrace{\left[\mathcal{L}\left(s\right)-\mathbf{d}_r\right]'\dot{\tilde{\mathbf{L}}}\left(s\right)\left(\mathbf{I}-\mathbf{F}\right)'\left(\mathbf{I}-\mathbf{F}\right)\tilde{\mathbf{L}}'\left(s\right)\left[\mathcal{L}\left(s\right)-\mathbf{d}_r\right]}_{\mathrm{WEMVA-like\ \ gradient}}. && (39)\\
&= \nabla J_{\mathbf{F}\mathrm{W}}+\nabla J_{\mathrm{F}\mathbf{W}} && (40)
\end{aligned}$$

When $\tau_{\mathrm{W}} \approx 0$ the first term in the gradient ($\nabla J_{\mathbf{F}\mathrm{W}}$ in equation 40) is close to the conventional FWI gradient because $\tilde{\mathbf{L}}\tilde{\mathbf{L}}'\left(\mathcal{L}\left(s\right)-\mathbf{d}_r\right) \approx \left(\mathcal{L}\left(s\right)-\mathbf{d}_r\right)$ by virtue of the approximation in equation 6. It also dominates the gradient, because the second term ($\nabla J_{\mathrm{F}\mathbf{W}}$ in equation 40) is small (it would be actually zero if we had not imposed the constraint of $c_0 = 0$).

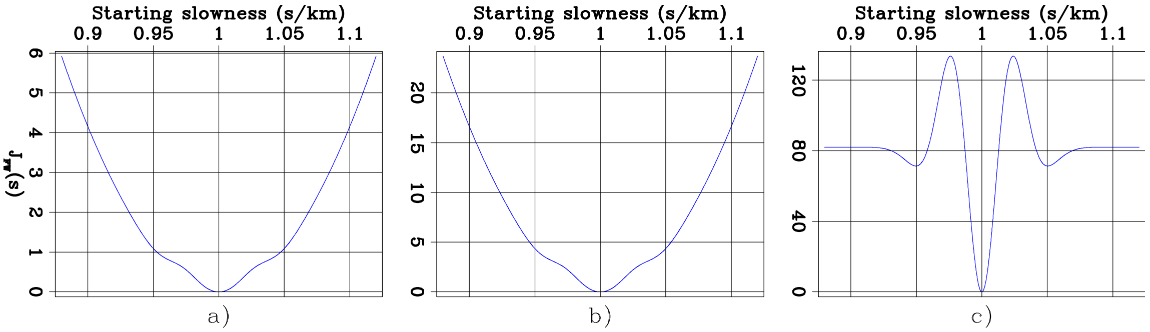
In contrast, when $\tau_{\mathrm{W}} \approx 1$, the first term in the gradient, $\nabla J_{\mathbf{F}\mathrm{W}}$, is small because the application of $\left(\mathbf{I}-\mathbf{F}\right)'\left(\mathbf{I}-\mathbf{F}\right)$ strongly attenuates the time lags in $\mathbf{c}$ that contribute the most to the backprojection of the residuals (see Figure 2b). Consequently the WEMVA-like term, $\nabla J_{\mathrm{F}\mathbf{W}}$, dominates the gradient, and ensures convergence towards the global minimum.

This behavior of the objective function in equation 37 is illustrated by Figure 14 through Figure 17. Figures 14 and 15 show $J_{\mathrm{FW}}\left(s_o\right)$ for $\mathbf{F} = \mathbf{F}_D$ and $\mathbf{F} = \mathbf{F}_G$, respectively. In each figure, the three panels correspond to different values of τ_{W}: for panels a) $\tau_{\mathrm{W}} = 1.0$, for panels b) $\tau_{\mathrm{W}} = 0.5$, and for panels c) $\tau_{\mathrm{W}} = 0.0$. The objective functions in the leftmost panels are convex. A gradient-based method would have no problems to converge towards the global minimum; the convergence, however, would be slow. In the middle panels, the objective functions are "tighter" but still convex. In contrast, the objective functions plotted in the rightmost panels are oscillatory and may cause similar convergence problems as experienced in the minimization of conventional FWI objective function. On the other hand, the high sensitivity of these oscillating objective functions to small changes in slowness may also enable the inversion to achieve high resolution, once we are close enough to the correct slowness. These observations suggest the application of an iterative inversion process that starts with wide focusing operators ($\tau_{\mathrm{W}} = 1.0$) and that slowly tightens the focusing operators toward $\tau_{\mathrm{W}} = 0.0$ as the data kinematics are fitted. Such an algorithm has the potential of achieving both robust global convergence from arbitrary starting model and fast local convergence close to the desired global minimum.

Figures 16 and 17 show the FWI-like gradient term for $\mathbf{F} = \mathbf{F}_D$ and $\mathbf{F} = \mathbf{F}_G$, whereas Figures 18 and 19 show the WEMVA-like gradient term for $\mathbf{F} = \mathbf{F}_D$ and $\mathbf{F} = \mathbf{F}_G$. Notice that the FWI-like gradient term is strongly oscillatory for all values of τ_{W}, but also that its amplitude is higher than the amplitude of the WEMVA-like term only for $\tau_{\mathrm{W}} = 0$ (panels c) in the figures. In contrast, the WEMVA-like term of the gradient that is shown in panels a) and b) is well-behaved for both choices of focusing operator. However, close to convergence; that is for s_o close to $\bar{s}$, the gradient is small. If we had to rely only on this gradient-component, the resolution of the inversion would be likely to suffer. In panels c) the WEMVA-like term becomes oscillatory and has "wrong" sign even close to convergence, but its amplitude insignificant compared to the amplitude of the corresponding FWI-like gradient terms.

We can also observe the WEMVA-like gradients in panels a) and b) are smoother when $\mathbf{F} = \mathbf{F}_G$ (Figure 19) than when $\mathbf{F} = \mathbf{F}_D$ (Figure 18). This difference may be indicative of a difference in robustness between the two focusing operators. To test this hypothesis we conducted a a similar test, but with a zero-phase wavelet in place of the Ricker-derived wavelet shown Figure 1a. The zero-phase wavelet has the same central frequency as the Ricker-derived wavelet, but it is more ringing. Consequently the objective functions and gradients are more oscillatory than the ones shown in previous figures.

Figure 20a shows the objective function computed using the highly-ringing zero-phase wavelet when $\mathbf{F} = \mathbf{F}_D$ and $\tau_{\mathrm{W}} = 1.0$; it corresponds to the objective function computed using the Ricker-derived wavelet and shown in Figure 14a. With this new wavelet, the objective function is not any more convex, and the total (FWI-like plus WEMVA-like term) gradient (Figure 20b) has two zero-crossing on each side of $\bar{s}$. On the contrary, the objective function computed with $\mathbf{F} = \mathbf{F}_G$ (Figure 21a) is still convex; its total gradient (Figure 21b) gets close to the horizontal axis, but it does not cross it, except at the expected zero-crossing at $s_o = \bar{s}$. This difference in behavior can be explained by comparing the WEMVA-like gradient terms (Figure 20c and Figure 21c). The one computed with $\mathbf{F} = \mathbf{F}_G$ is smoother than the one computed with $\mathbf{F} = \mathbf{F}_D$.



a) b) c)

Figure 14: $J_{\mathrm{FW}}(s_o)$ computed with $\mathbf{F} = \mathbf{F}_D$ and with: a) $\tau_{\mathrm{W}} = 1.0$, b) $\tau_{\mathrm{W}} = 0.5$, and c) $\tau_{\mathrm{W}} = 0.0$. [**ER**] biondo1/. J-W-F-DSO

DISCUSSIONS

All the three approaches that we presented to solve the extended inverse problem show promises to lead to inversion algorithms with robust global convergence. However, we have

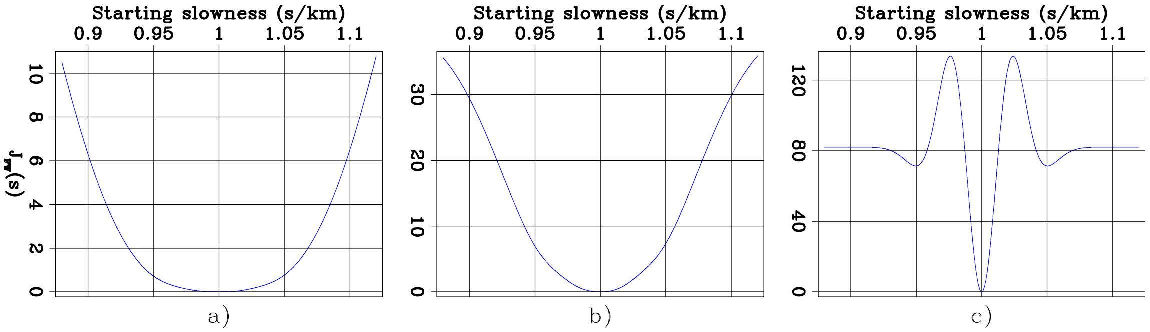


Figure 15: $J_{\mathrm{FW}}\left(s_o\right)$ computed with $\mathbf{F} = \mathbf{F}_G$ and with: a) $\tau_{\mathrm{W}} = 1.0$, b) $\tau_{\mathrm{W}} = 0.5$, and c) $\tau_{\mathrm{W}} = 0.0$. [**ER**] biondo1/. J-W-F-Band

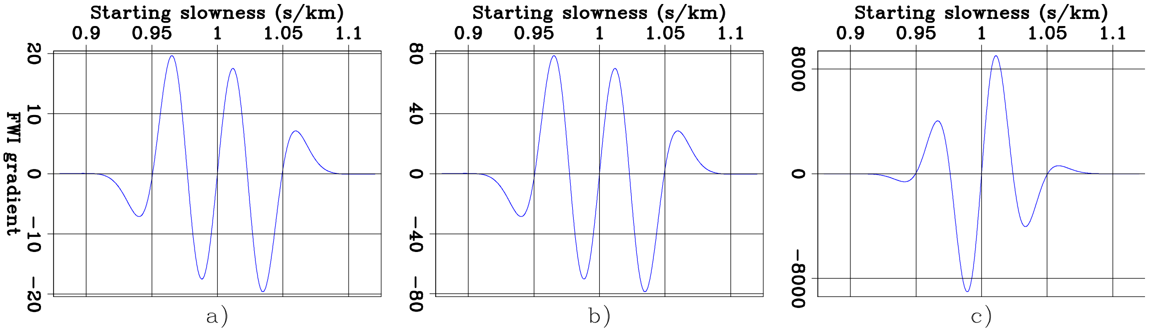


Figure 16: $\nabla J_{\mathbf{FW}}\left(s_o\right)$ computed with $\mathbf{F} = \mathbf{F}_D$ and with: a) $\tau_{\mathrm{W}} = 1.0$, b) $\tau_{\mathrm{W}} = 0.5$, and c) $\tau_{\mathrm{W}} = 0.0$. [**ER**] biondo1/. Grad-F-DSO

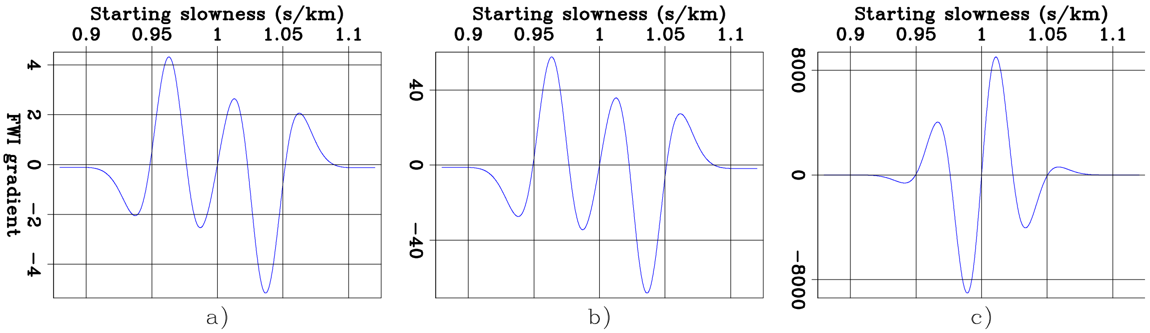
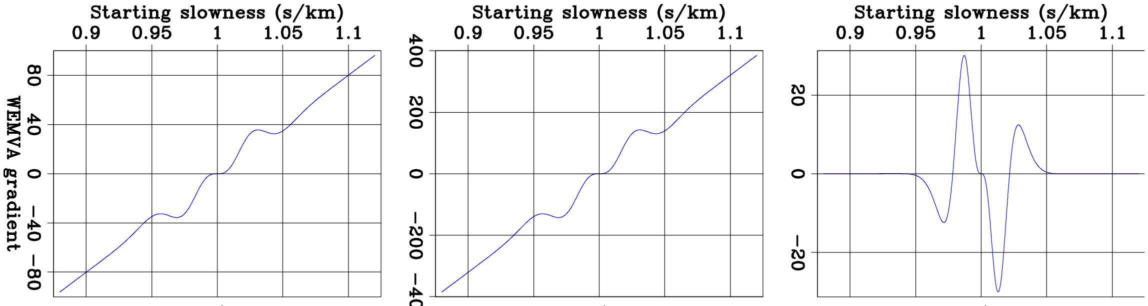
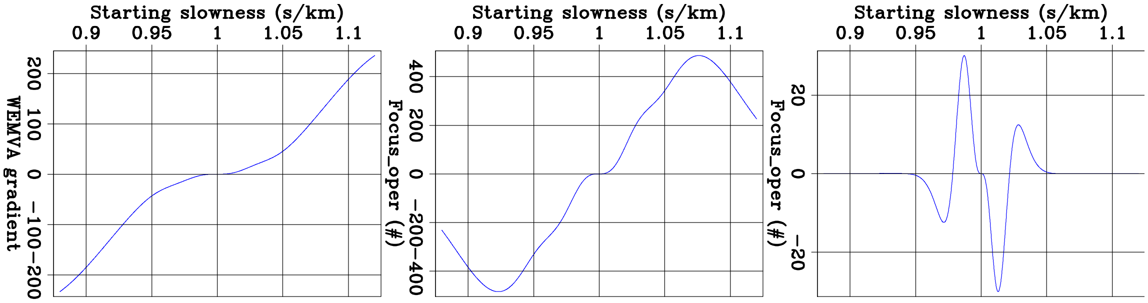


Figure 17: $\nabla J_{\mathbf{FW}}\left(s_o\right)$ computed with $\mathbf{F} = \mathbf{F}_G$ and with: a) $\tau_{\mathrm{W}} = 1.0$, b) $\tau_{\mathrm{W}} = 0.5$, and c) $\tau_{\mathrm{W}} = 0.0$. [**ER**] biondo1/. Grad-F-Band



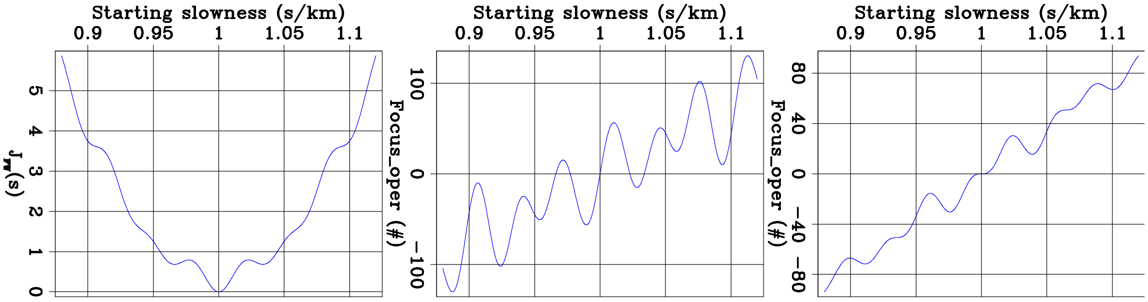
a) b) c)

Figure 18: $\nabla J_{\mathrm{F}\mathbf{W}}\,(s_o)$ computed with $\mathbf{F} = \mathbf{F}_D$ and with: a) $\tau_{\mathrm{W}} = 1.0$, b) $\tau_{\mathrm{W}} = 0.5$, and c) $\tau_{\mathrm{W}} = 0.0$. **[ER]** biondo1/. Grad-W-DSO



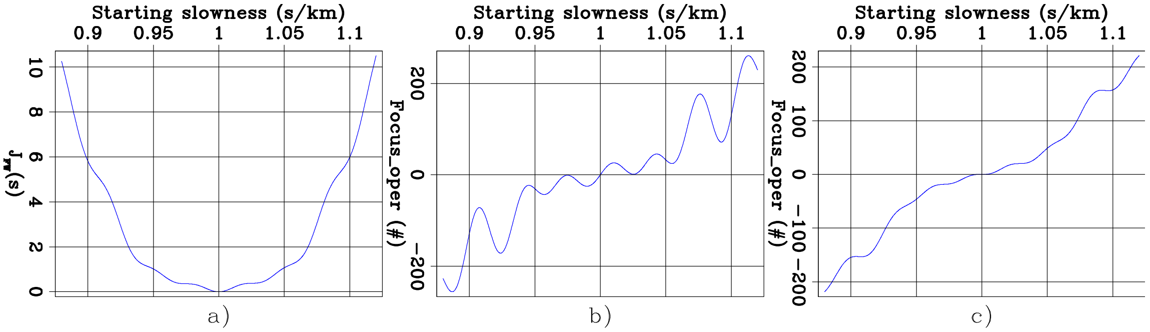
a) b) c)

Figure 19: $\nabla J_{\mathrm{F}\mathbf{W}}\,(s_o)$ computed with $\mathbf{F} = \mathbf{F}_G$ and with: a) $\tau_{\mathrm{W}} = 1.0$, b) $\tau_{\mathrm{W}} = 0.5$, and c) $\tau_{\mathrm{W}} = 0.0$. **[ER]** biondo1/. Grad-W-Band



a) b) c)

Figure 20: a)$J_{\mathrm{FW}}\,(s_o)$ computed with a zero-phase wavelet that was more ringing than the Ricker-derived wavelet used for the previous figures. b) $\nabla J_{\mathrm{FW}}\,(s_o)$, and c) $\nabla J_{\mathrm{F}\mathbf{W}}\,(s_o)$. All these three curves were computed with $\mathbf{F} = \mathbf{F}_D$ and $\tau_{\mathrm{W}} = 1.0$. **[ER]** biondo1/. J-Grad-DSO-bp



a) b) c)

Figure 21: a)$J_{\mathrm{FW}}\left(s_o\right)$ computed with a zero-phase wavelet that was more ringing than the Ricker-derived wavelet used for the previous figures. b) $\nabla J_{\mathrm{FW}}\left(s_o\right)$, and c) $\nabla J_{\mathbf{FW}}\left(s_o\right)$. All these three curves were computed with $\mathbf{F} = \mathbf{F}_G$ and $\tau_{\mathrm{W}} = 1.0$. **[ER]** biondo1/. J-Grad-Band-bp

not sufficiently developed our analysis to draw firm conclusions on the rate of convergence of the proposed methods. Since the development of efficient inversion algorithms is one of our main goals, further work in this direction is needed.

The 1D wave-propagation problem we used to analyze the proposed inversion methods has two useful advantages: 1) it enables comprehensive analysis of global convergence because it is computationally fast and 2) objective functions and gradients can be analyzed as simple 1D plots. However, it has also two (related to each other) main shortcomings: 1) it models transmitted events but not reflected ones, and 2) its model space (a simple scalar) cannot be decomposed into different scales, and thus does not enable insights on how different model scales (long vs. short wavelengths) behave during the inversion process.

REFERENCES

Almomin, A. and B. Biondi, 2014, Preconditioned tomographic full waveform inversion by wavelength continuation: SEG Technical Program Expanded Abstracts, **33**, 944–948.

Biondi, B. and A. Almomin, 2014, Simultaneous inversion of full data bandwidth by tomographic full waveform inversion: Geophysics, **79**, WA129–WA140.

Biondi, B. and P. Sava, 1999, Wave-equation migration velocity analysis: SEG Technical Program Expanded Abstracts, **18**, 1723–1726.

Sava, P. and B. Biondi, 2004, Wave-equation migration velocity analysis-I: Theory: Geophysical Prospecting, **52**, 593–606.

Shen, P. and W. W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, VE49–VE59.

Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790.

Symes, W. W. and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663.

Zhang, Y. and B. Biondi, 2013, Moveout-based wave-equation migration velocity analysis: Geophysics, **78**, U31–U39.

Stanford Exploration Project, SEP165, October 8, 2016

Deblending using the radiality attribute

Joseph Jennings and Shuki Ronen

ABSTRACT

The non-uniqueness of simultaneous source deblending inversion is a challenge in simultaneous source separation. We propose to add another constraint to this inversion using radiality, an attribute that can be computed from multicomponent data that are commonly recorded during ocean-bottom node (OBN) or multi-sensor streamer surveys. We describe a simple, proof-of-concept scheme that demonstrates the use of radiality as an additional constraint and show results on an OBN field dataset.

INTRODUCTION

Simultaneous source shooting is an emerging method used to reduce acquisition costs while still providing high quality seismic data. Additionally, it is used to improve data quality while keeping the cost constant. In order to accurately image these data, there exist many challenges due to the cross-talk from interfering sources. The process of separating the data and removing the seismic interference between the simultaneous sources is called deblending.

One proposed solution for deblending is to separate the interfering source from the dominant source via a sparse inversion approach (Abma et al., 2010; Ayeni et al., 2011). Intrinsically, this approach poses a highly-underdetermined problem that must be constrained via a regularization term in addition to the data fitting term in the objective function. Effective choices for this regularization have been using continuity in the common-receiver domain.

We propose an additional constraint for deblending inversions from using multicomponent data. We show that using additional information from the other components of an ocean-bottom node (OBN) dataset can provide a better starting model for deblending inversion. As of now, we are using four component OBN data. However, this method is applicable to three component multi-sensor streamer data and perhaps also with attributes calculated from single- or dual-component conventional streamer data with the in-line horizontal component computed from the in-line spatial derivative of the data and the cross-line horizontal component unknown.

In Jennings and Ronen (2016), we described both radiality and source similarity attributes. In this report, we further develop the radiality attribute. We first describe how using the radiality attribute can provide additional information that can aid deblending inversions and show examples on a synthetically blended field dataset from the North Sea (Alves, 2015). We then provide a simple proof-of-concept method that shows that using radiality, we can create more accurate starting models for a deblending inversion and we demonstrate these results on the same North Sea dataset.

METHODOLOGY

Additional information from multicomponent data

To understand how using multicomponent data can aid in deblending seismic data, let us consider a simple example. Figure 1 shows the acquisition geometry of a synthetically blended dataset from the Forties oil field in the North Sea. Note that because these data were blended synthetically, this creates a worst case scenario in terms of signal to noise ratio.

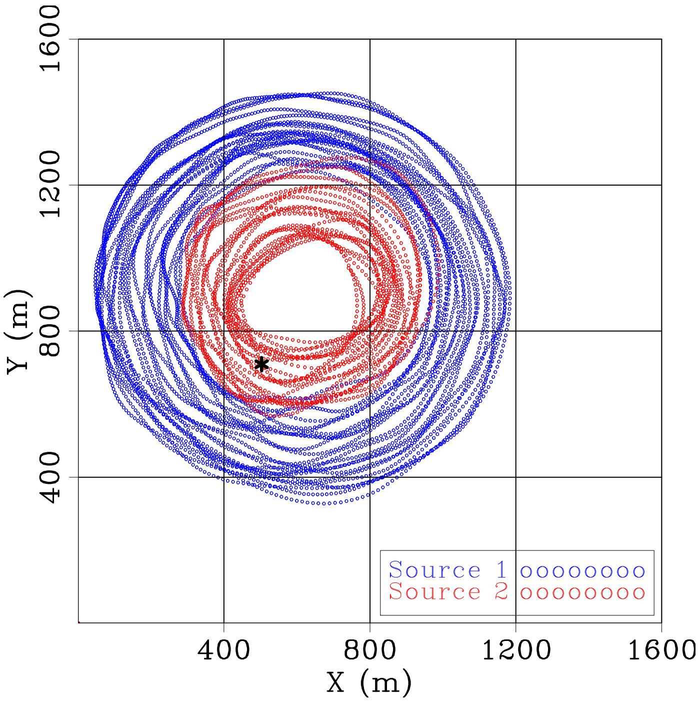


Figure 1: The simultaneous source acquisition geometry over the Forties oil field in the North Sea. Only two sources were used in the blending of these data and their locations are indicated by the red and blue circles. Both sources circled around a platform under which ocean-bottom nodes were positioned. The black star indicates the location of the receiver node use in this study. [**ER**] joseph1/. src-reg-002

For the purpose of explaining the radiality attribute, let us consider only two shots that have a source-receiver azimuthal difference of approximately 90°. These two source positions and the node location are shown in Figure 2. The corresponding multicomponent data recorded at the node location from these two shots are shown in Figure 3. Examining the hydrophone and vertical components (Figures 3(a) and 3(b)), we can clearly observe both source 2 and source 1 as expected from a trace of the blended common receiver gather. In Figure 3(c), we observe the radial trace rotated toward the azimuth of source 1. It is apparent that after rotation, source 1 is the dominant source on the radial component and source 2 has much less energy. Likewise, on the transverse component (Figure 3(d)), which points almost directly toward the source 2 azimuth, source 2 is the dominant source and source 1 has much less energy. This is because almost all the waves that radiate from any acoustic source have radial polarity.

We observe a similar result when the data are rotated toward the azimuth of source 2. Figure 4 shows the result of rotating the data toward the azimuth of source 2. Again, observing the radial and transverse components (Figures 3(c) and 3(d)), it is clear that by rotating into the azimuth of source 2, we can identify the dominant source at any time on a blended trace.

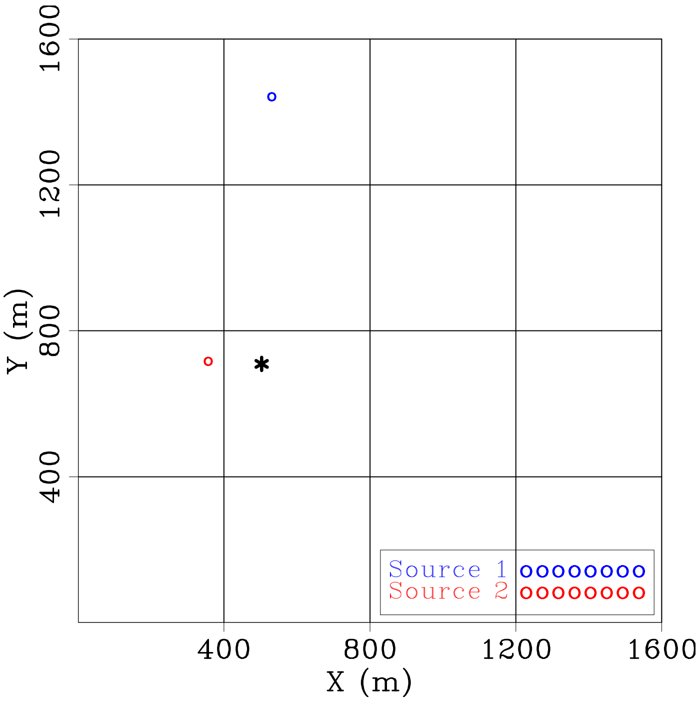
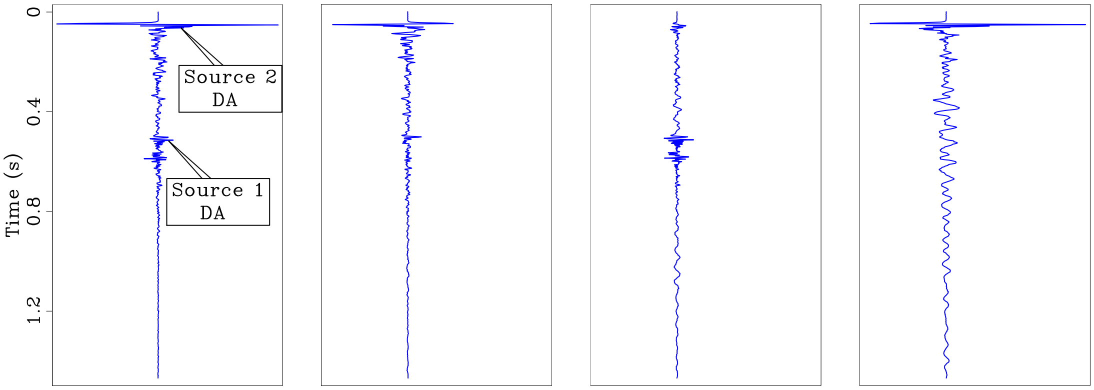


Figure 2: Two source locations selected from the locations shown in Figure 1. These locations were chosen as they have an approximate source-receiver azimuthal difference of 90°. [**ER**] joseph1/. shw-azim90-002



(a) (b) (c) (d)

Figure 3: Data recorded from source locations shown in Figure 2. 'DA' indicates direct arrival. Source two was recorded first and then source one. Also note that source two is much stronger than source one in amplitude due to its proximity to the node. Panel (a) shows the hydrophone component, (b) the vertical component, (c) the radial component rotated toward the azimuth of source 1, and (d) the transverse component (rotated perpendicular to the azimuth of source 1). Note how source 1 is distinct in (c) and source 2 in (d). [**ER**] joseph1/. b12-h-raw877-anno, b12-v-raw877, b12-r-raw877, b12-t-raw877

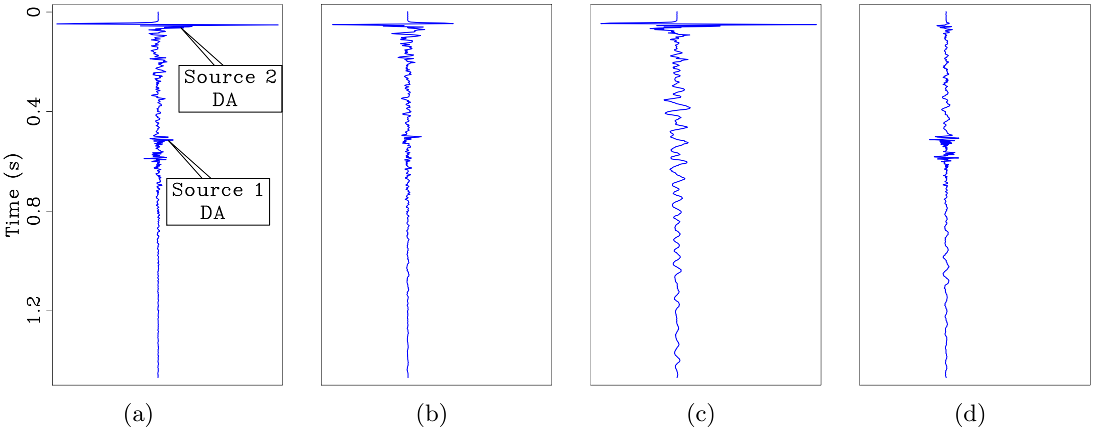


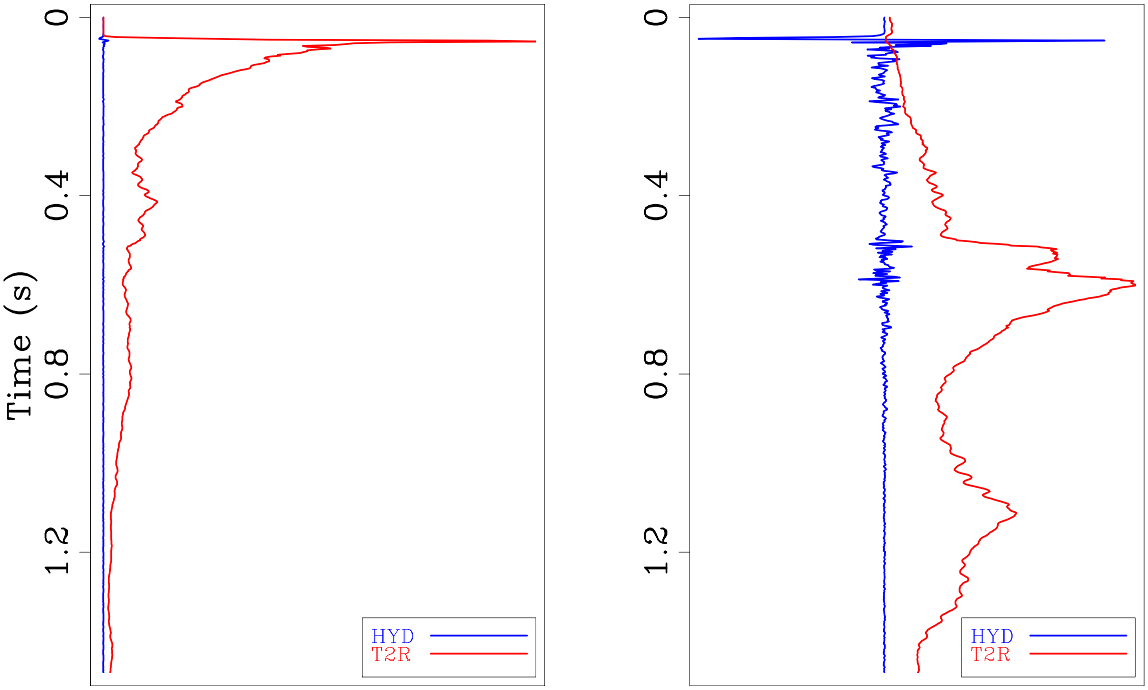
Figure 4: Same as data shown in Figure 3 but now rotated into the azimuth of source 2. As in Figure 3, 'DA' indicates the direct arrival. Panel (a) shows the hydrophone component, (b) the vertical component, (c) the radial component rotated toward the azimuth of source 2 and (d) the transverse component (rotated perpendicular to the azimuth of source 2) Note how source 2 is distinct in (c) and source 1 in (d). [**ER**] joseph1/. b21-h-raw624-anno, b21-v-raw624, b21-r-raw624, b21-t-raw624

Radiality and deblending

To turn this identification capability into a deblending method, we produce a radiality attribute which may be used as a constraint in an inversion problem. Using radiality, we can detect the presence of the dominant source at any time on a blended trace from a multicomponent common receiver gather. This information stems from the fact that when rotating into the different source-receiver azimuths, one source is significantly more radial than the other relative to the location of the node. Therefore, we desire to quantify how radial the data are at any given time. In Jennings and Ronen (2016), we defined the radiality of a trace as the ratio of the radial and transverse envelopes. Thus, the radiality is small when non-radial interfering sources are recorded. As an extension of radiality, we define a new attribute as the ratio of the transverse envelope and the radiality. Mathematically, this may be written as:

$$\frac{E_{\text{transverse}}}{\text{Radiality}} = \frac{E_{\text{transverse}}}{\frac{E_{\text{radial}}}{E_{\text{transverse}}}} = \frac{E^2_{\text{transverse}}}{E_{\text{radial}}},$$

which is just the ratio of the square of the transverse envelope and the radial envelope. We denote this new radiality attribute as $\frac{T^2}{R}$. Figure 5 shows the overlay of this attribute over the hydrophone component. From this figure, it is clear that the T^2/R attribute detects the interfering source relative to the dominant source. Note that from this point onward when referring to the T^2/R attribute, we use the term "radiality attribute".



(a) (b)

Figure 5: Radiality attribute overlain on the hydrophone component. The radiality attribute of source 1 is shown in panel (a) and the radiality attribute from source 2 is shown in panel (b). In panel (a) the radiality attribute is maximum over the direct arrival of source 2, the interfering source. Likewise, in panel (b), the radiality attribute is maximum over source 1, the interfering source in this panel. [**ER**] joseph1/. t2r-h-comp1-002b12-877,t2r-h-comp1-002b21-624

RESULTS

As a simple proof-of-concept of how this radiality attribute can be used in a deblending inversion, we use it to mute the data from the interfering source and then interpolate to fill in the gaps. The interpolated data give us a better starting model for a deblending inversion. Figure 6 shows 1000 traces of the blended hydrophone component of a common receiver gather from the same North Sea, Forties oil field dataset described previously. The computed radiality attribute on this common receiver gather is shown in Figure 7(a). The oscillating light and dark regions are indicative of the azimuthal difference of the sources (i.e. brighter regions indicate an azimuthal difference closer to 90°). Figure 7(b) shows the deblended data after muting the interference and linearly interpolating from shot to shot.

Comparing Figures 6 and 7(b) it is clear that in the regions of approximately 15-90°azimuthal difference between the sources, much of the interference was removed. However, in areas where this is not the case, the interference is still largely present. Figures 8(a)-8(c) show a zoom-in of each step of the deblending scheme. Again, from these figures it is clear that the radiality attribute has successfully detected much of the interference in regions of large azimuthal difference. In regions where the azimuthal difference is small, the interference has been largely untouched.

As we only linearly interpolated from trace to trace, in regions where there exist steep events or large gaps, the linear interpolation will not be very accurate. If we first flatten the data with a normal moveout correction (NMO), and interpolate the flattened gathers, we achieve better interpolated data. The results of interpolating in the NMO domain are shown in Figure 8(d). The difference between this result and the original unblended data (shown in Figure 8(e)) is shown in Figure 8(f). In this figure, it is clear that much of the interference has been removed. Note also the errors in interpolation near the direct arrival (between the shot range 200-250 and 500-600).

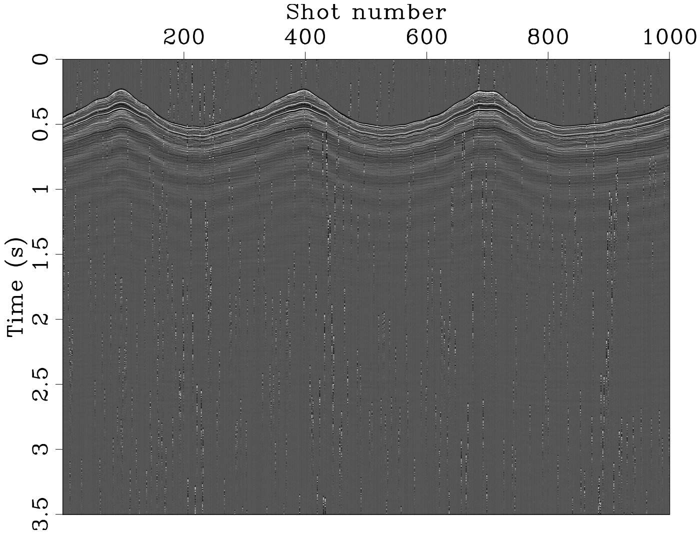
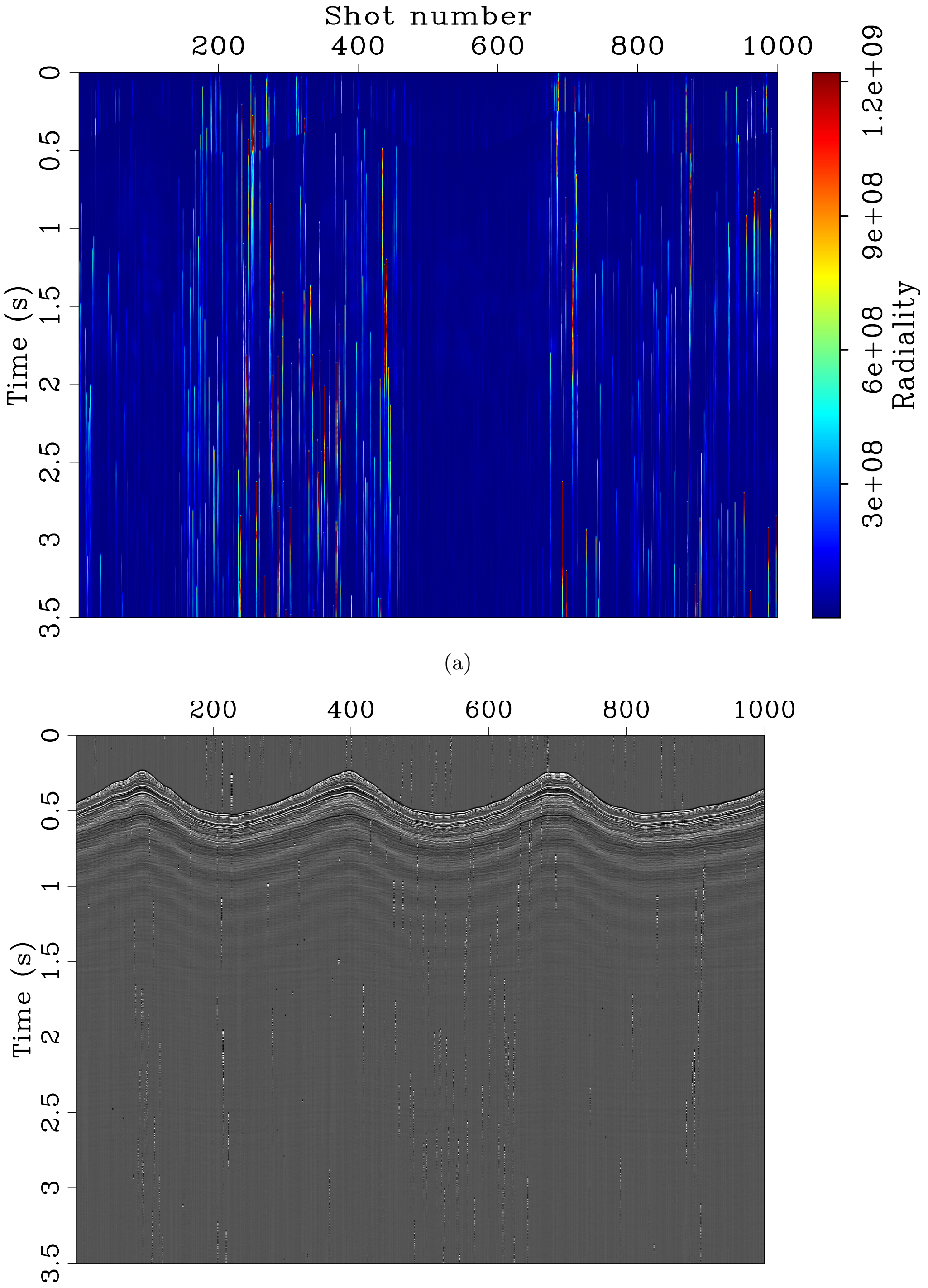


Figure 6: 1000 traces from the hydrophone component of a blended common receiver gather from the Forties oil field data. **[ER]** joseph1/. hydro-002b12

DISCUSSION

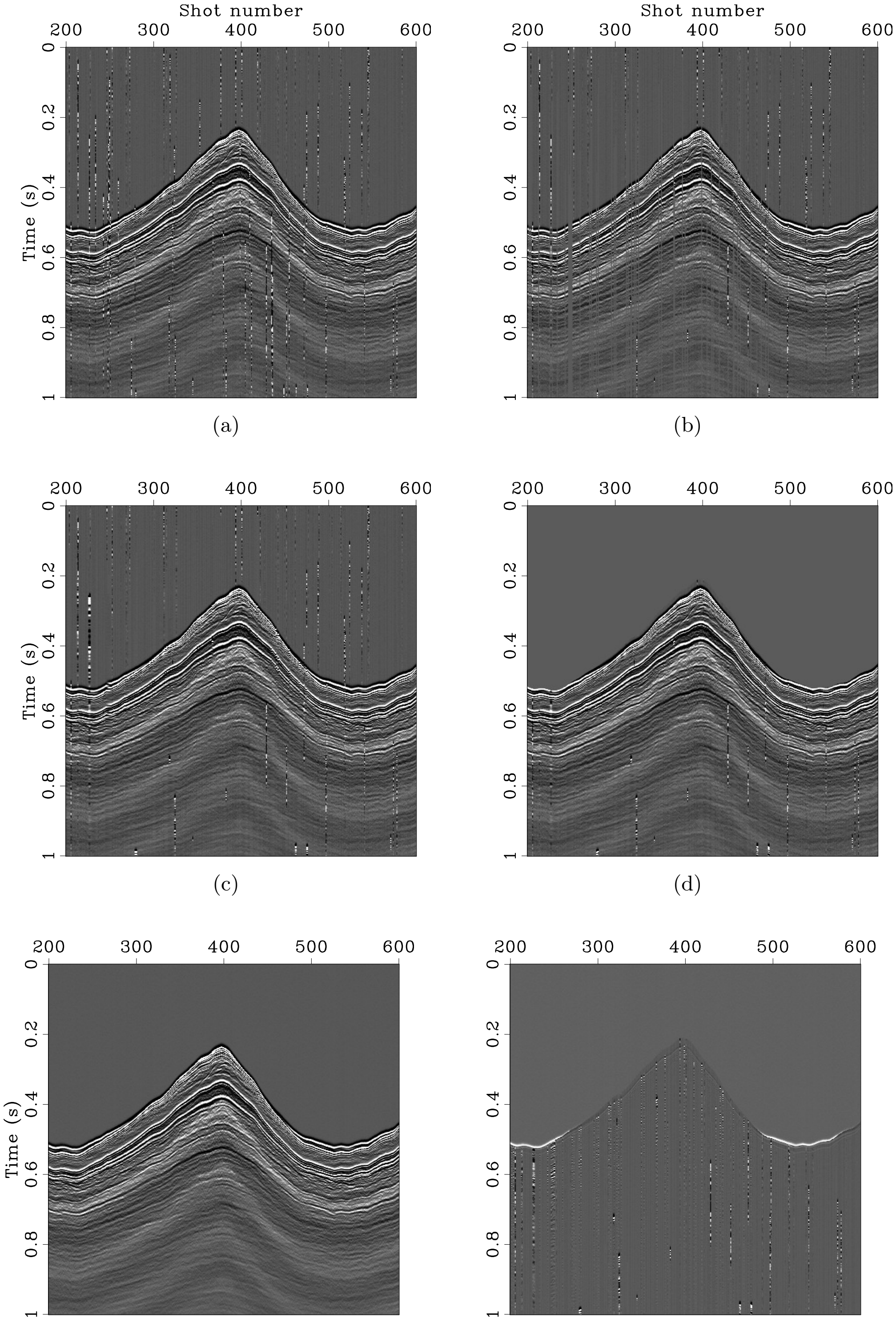
One limitation of this method, as shown in Figure 8, is that if the interfering source is in the same azimuth as the primary source, or 180° away, the radiality attribute is of little use.



(a)

(b)

Figure 7: (a) The radiality attribute computed over the 1000 traces shown in Figure 6. (b) The deblended data obtained via muting the interfering shots and then linearly interpolating from shot to shot. Note in panel (a) that the oscillating dark and light regions occur due to the changing source azimuths. From panel (b) it is clear that where the radiality attribute is high (the regions where the azimuthal difference between the sources is approximately 15-90°), our deblending scheme removes the interfering sources with much greater success. [**ER**] joseph1/. t2or-002b12, interped-002b12



(a) (b)

(c) (d)

(e) (f)

Figure 8: A zoom-in of the common receiver gather from shot 200 to 600. (a) The synthetically blended data. (b) The data after muting the interfering shots using the radiality attribute (Figure 7(a)). (c) The muted data linearly interpolated from shot to shot. (d) The muted data linearly interpolated from shot to shot in the NMO domain. (e) The original unblended data included for comparison and (f) the difference between the deblended data shown in (d) and the unblended data (e). [**ER**]
joseph1/. h-b12-raw,mut-b12-h,int-b12-h,nmo-b12-h,h-ub1-raw,diff-b12-h

While it is possible to extend the method to 180° separation, it is inherently impossible to extend it to 0° separation. Another limitation, which was not apparent in the deblending demonstration of Figure 8 but is a potential problem, is related to the synchronization of the sources. If the sources are synchronized, then filling in the declared missing data with interpolation will be impossible as the gaps will extend laterally from one trace to another on a common receiver gather. Both of these limitations can be mitigated by survey design to minimize the shots with small azimuthal difference and to avoid synchronization between the sources. For example, avoiding synchronization could be done if each source vessel moved at a slightly different velocity. Mitigated or not, with these limitations, radiality provides additional information and serves as an additional constraint for deblending inversions.

Another limitation of the method is that primary source waves may arrive on a non-radial direction if they are reflected or scattered from the side. Using radiality, such waves may erroneously be identified as seismic interference. For this reason, in the future it will be important to develop a more sophisticated constraint on the seismic interference using a probability instead of a harsh mute as we have shown so far.

CONCLUSION

We have shown that using radiality, an attribute computed from the horizontal components of the geophone or ocean-bottom node, we can identify the presence of the dominant source at anytime given that the source vessels have different source-receiver azimuths. We provided a simple proof-of-concept example in which using the radiality attribute, we zeroed the energy from the interfering source. We then interpolated the zeroed data which provided a result that has less interference than the blended data and thus could prove to be a more accurate starting model as input to a full deblending inversion.

ACKNOWLEDGEMENTS

The authors would like to thank Apache Corporation for the permission to show the data and acknowledge the sponsors of the SEP for their financial and intellectual support.

REFERENCES

Abma, R., T. Manning, M. Tanis, J. Yu, and M. Foster, 2010, High quality separation of simultaneous sources by sparse inversion: Presented at the 72nd EAGE Conference and Exhibition incorporating SPE EUROPEC 2010.

Alves, G., 2015, Overview of the apache forties data set: SEP-Report, **160**, 175–180.

Ayeni, G., A. Almomin, D. Nichols, et al., 2011, On the separation of simultaneous-source data by inversion: Presented at the 2011 SEG Annual Meeting.

Jennings, J. and S. Ronen, 2016, Separation of simultaneous source blended data using radiality and source similarity attributes: SEP-Report, **163**, 239–256.

Stanford Exploration Project, SEP165, October 8, 2016

Detecting karst caverns by pattern recognition

Fantine Huot and Robert Clapp

ABSTRACT

Strong localized heterogeneities in the subsurface, such as karst caverns and sinkholes, cause scattering of seismic waves, thereby degrading the images obtained in conventional processing. We explore the possibility of using pattern recognition techniques for detecting these strong heterogeneities from seismic data. Through synthetic models, we generate data with significant scattering. We then perform reverse time migration (RTM) and use various pre-processing techniques to engineer features fit for supervised learning algorithms. Eventually, we use support vector machines (SVM) to classify these features and retrieve the approximate cavern locations.

INTRODUCTION

The Tengiz carbonate platform in northwestern Kazakhstan is one of the largest producing oil fields in the world. Recently, exploration has targeted karst-like zones with cavernous porosity along the margin of the platform. Lester et al. (2015) showed that these karst caverns appear as localized high-amplitude events on seismic volumes but can also resemble residual noise that may have persisted through processing and imaging. Such localized features induce positional uncertainty in the migrated velocity model and can represent drilling hazards. While various methodologies such as diffraction migration or beam migration (Fomel et al., 2007; Berkovitch et al., 2009; Lester et al., 2015) have been proposed to address the issue of imaging these strong heterogeneities, herein we investigate the potential of techniques commonly used in pattern recognition.

The first algorithm for pattern recognition was introduced 80 years ago (Fisher, 1936). With the advent of computers and the information age, statistical learning has become a highly explored field in many scientific areas as well as marketing, finance, and other business disciplines. In recent years, new and improved software packages have significantly eased the implementation burden for many statistical learning methods, providing scientists and practitioners with complete toolkits for training, testing, and deploying models with well-documented examples for all these tasks (Collobert et al., 2002; Pedregosa et al., 2011; James et al., 2013; Jia et al., 2014). With algorithms automatically tracking faces in photographs (Osuna et al., 1997), what would prevent us from training machines to detect specific seismic responses in our data?

We first generate synthetic seismic data from a cavern model. We then perform reverse time migration (RTM) and apply various pre-processing techniques, such as continuous wavelet transforms (CWT) and principal component analysis (PCA), to build appropriate input features for our pattern recognition problem. We then train a support vector machine (SVM) classifier to detect the migrated seismic signature associated with caverns and apply this classifier to different portions of the data to retrieve the approximate cavern locations.

SYNTHETIC DATA GENERATION

Cavern model

The methodology used in this study is based on the one presented by Huot and Clapp (2016). We start by generating seismic data from a synthetic cavern model.

For this purpose, we create a three-dimensional synthetic model of an underground karst channel system in a limestone bedding as illustrated in Figure 1. Caverns of different scales are inserted at random locations in the model and connected with channels partially filled with water. In the following, we use four 2D slices along the y-axis to generate data which we use to train our machine learning model, and a 2D slice along the x-axis to generate data to test our classification performance.

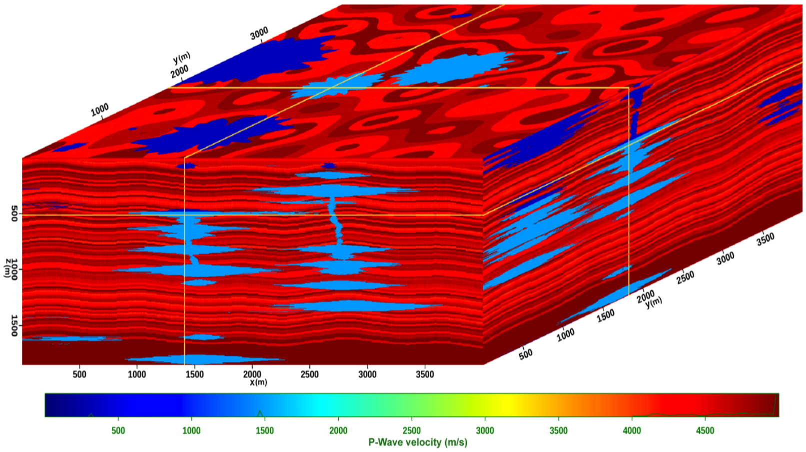


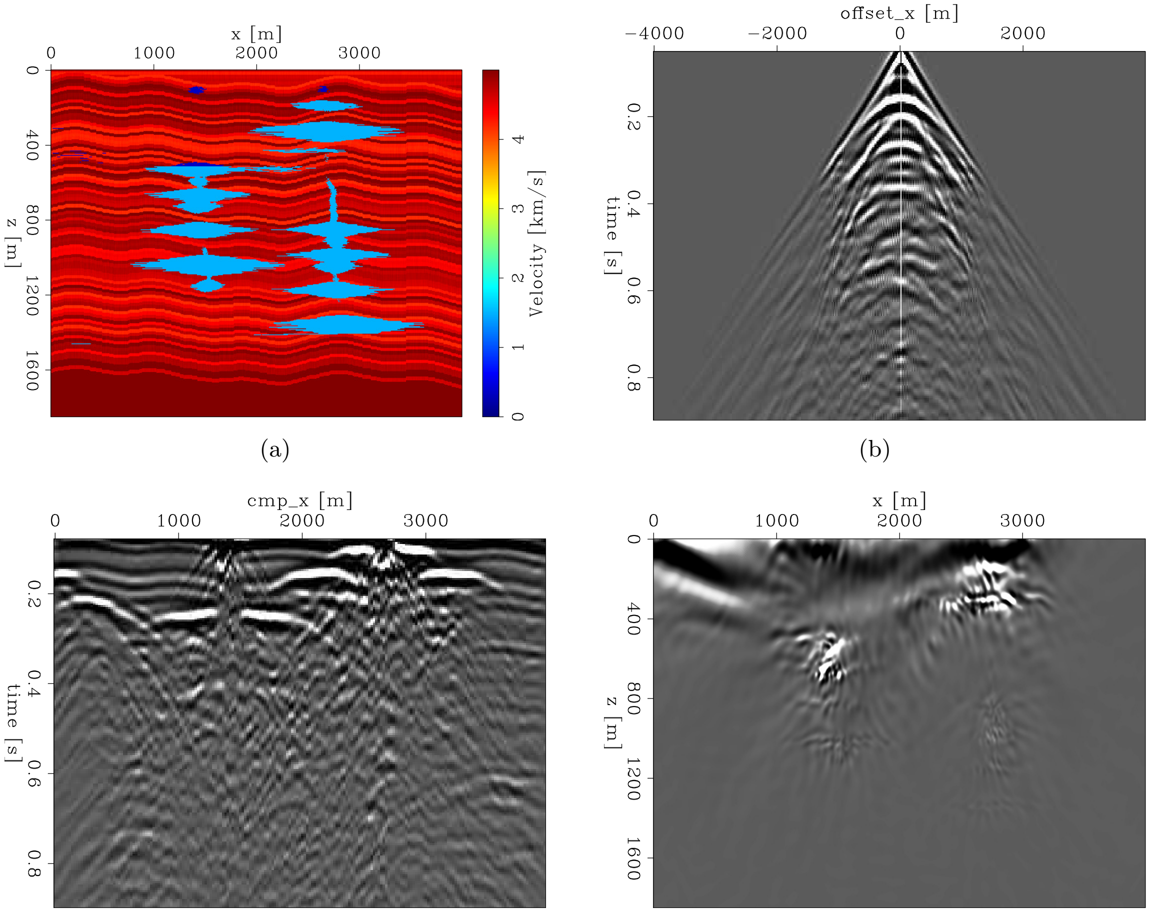
Figure 1: 3D synthetic cavern model built for this study. It features a randomly generated karst channel system in a limestone bedding. The model is color-coded by P-wave velocity. The detailed description of this model is provided in Huot and Clapp (2016). [**NR**] fantine1/. model

Data modeling

For each of these 2D model slices, we generate synthetic seismic data using wave propagation code based on the one developed by Alves (2015). We use a Ricker-type explosive source with a peak frequency of 20 Hz. The recordings for a single shot clearly illustrate the scattering effect associated with the presence of the caverns (Figure 2(b)). From the zero-offset common midpoint (CMP) gathers, we observe that the caverns incoherently scatter the reflected energy (Figure 2(c)), thereby preventing accurate identification of the cavern locations.

BUILDING FEATURES FOR MACHINE LEARNING

We now process the generated data to build features with a suitable format for machine learning algorithms.



(a) (b)

(c) (d)

Figure 2: (a) One of the 2D model slices from which we generate data to train our classification algorithm. (b) Single shot gather obtained from this model slice with a source located at the surface at x = 2000 m. (c) Zero-offset common midpoint gather. (d) RTM image generated from this model slice. We clearly distinguish the scattering effect due to the presence of the caverns. [**CR**] fantine1/. train,shot,cmp,rtm

To set up a classification problem, we need to associate each portion of the data with a binary label indicating whether or not it corresponds to a cavern location. Therefore, we first perform reverse time migration (RTM) in order to map the data to model space, allowing us to label the migrated image points using the synthetic model slices as reference. Figure 2(d) provides an example of one of these RTM images. The RTM image also suffers from the strong scattering due to the presence of the caverns.

Starting with the labeled RTM images, we then apply various processing steps to compute multiple data features:

- We perform both acoustic and elastic modeling and migration, creating two sets of features.
- We apply a Laplacian filter on the RTM images to attenuate the migration artifacts (Biondi, 2006).
- We apply gain power with depth to compensate for amplitude attenuation.
- We apply Gaussian smoothing and spatial frequency bandpassing to reduce high frequency jitter.
- We apply continuous wavelet transforms (CWT). CWT are commonly used in pattern recognition, as they have the ability to decompose complex patterns into elementary forms. They measure the similarity between a signal and an analyzing wavelet by comparing the input signal to shifted and compressed or stretched versions of the wavelet. An overview of CWT is provided in Huot and Clapp (2016). In this study, we use both Ricker and Morlet wavelets as the mother wavelets, producing multiple sets of features. To each image point, we associate the full panel of dyadic frequencies obtained after applying CWT, as described by Huot and Clapp (2016).
- We run a sliding window of three different sizes ($8\,\mathrm{m} \times 8\,\mathrm{m}$, $12\,\mathrm{m} \times 12\,\mathrm{m}$, $16\,\mathrm{m} \times 16\,\mathrm{m}$) over the image. We associate the full set of features captured by each sliding window to the image point located at its center.
- We standardize all the variables to have zero mean and standard deviation one.
- For faster computation, we perform principal component analysis (PCA). When faced with a large set of correlated variables, principal components allow us to summarize the dataset with a smaller number of representative variables that collectively explain most of the variability in the original set. A full overview of PCA is provided in Huot (2016). Herein, for each feature set, we use the minimum number of principal components that collectively explain at least 90% of the total variance.

Using different combinations of these processing steps, we build multiple sets of features and data characteristics associated with binary labels. We obtain distinct sets of training features and testing features using the data generated from the different 2D model slices.

CLASSIFICATION

Support vector machines (SVM)

In the following, we set up a classification algorithm to retrieve the cavern locations from the data features we have designed. As our problem contains highly correlated features, we decide to use support vector machines (SVM), known to be efficient for these type of problems (Hsu et al., 2003; Hastie et al., 2005).

SVM classifiers use large sets of labeled training data to build a decision function. They can then be applied on other portions of data to predict the corresponding labels. SVM classifiers are effective in high dimensional spaces, and use only a subset of training points in the decision function, making them memory efficient. They are versatile as many different kernel functions can be specified for the decision function, making it possible to define non-linear boundaries. An overview of the theory behind SVM is provided in Huot and Clapp (2016).

Implementation

In recent years, new and improved software packages have significantly eased the implementation burden for many statistical learning methods. In this study, we use the following Python packages:

- For data visualization and performing operations on data: pandas (http://pandas.pydata.org/)
- For machine learning models: scikit-learn (http://scikit-learn.org/stable/)
- For plotting data: matplotlib (http://matplotlib.org/)
- For progress monitoring: tqdm (https://pypi.python.org/pypi/tqdm)

Evaluation score

To evaluate how well our SVM classifier retrieves the cavern locations, we have to introduce an evaluation score. Common metrics used for binary classification are precision, recall and F1-score (Hastie et al., 2005). A visual representation of these metrics is provided in Figure 3. Precision is the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved:

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$$

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$$

The F1-score is the weighted harmonic mean of precision and recall, and hence provides a combined measure:

$$F = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

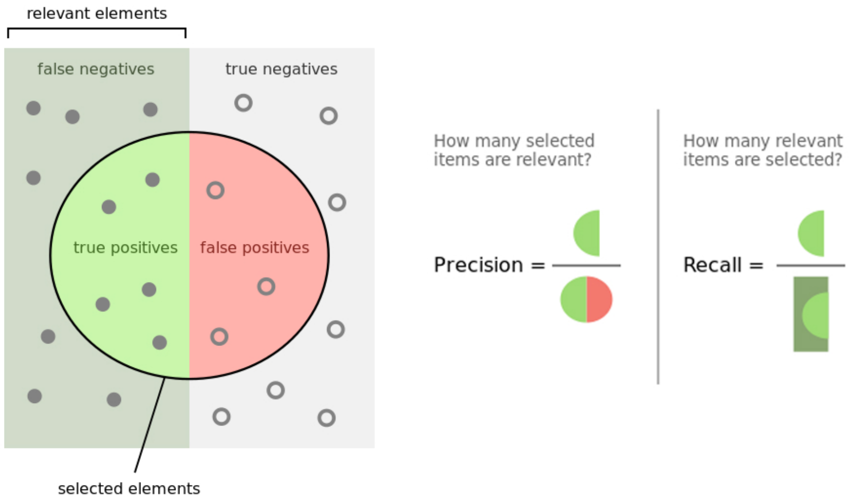


Figure 3: Visual representation of common metrics used for binary classification: precision and recall. Source: en.wikipedia.org/wiki/Precision_and_recall/media/File:Precisionrecall.svg [**NR**] fantine1/. metric

Classification results

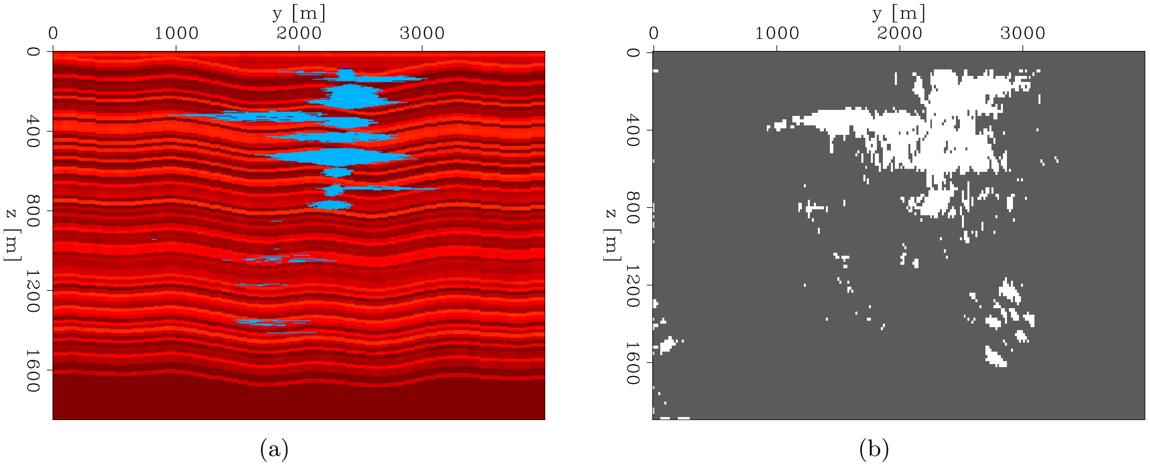
We build an SVM classifier for each different feature set. We use a radial basis function (RBF) kernel. For each classifier, we select the SVM cost parameter C that provides best evaluation score among 7 different values : 0.001, 0.01, 0.1, 1, 10, 100, 1000. A detailed description of these parameters is provided in Huot and Clapp (2016).

The feature sets that yield best classification results are those that have CWT in the pre-processing steps. The sliding window also improves classification results significantly. However, even the best classification results we obtained return many false positives, as illustrated in Figure 4. The associated performance metrics are provided below:

	Precision	Recall	F1-score
Limestone bedding	0.99	0.95	0.97
Caverns	0.40	0.80	0.53
Avg / total	0.97	0.94	0.95

When examining the classification results, it appears that the classifier also returns a certain number of false positives on the training data, indicating that our classification model suffers from high bias. This problem can potentially be improved by adding more features. Instead of testing sets of features one-by-one to identify which features yield best evaluation score, it may be preferable to provide the classifier with the full set features that can be computed from all the pre-processing steps, and skip the PCA.

It also appears that the test error decreases with the number of 2D model slices used for the training data, suggesting that a larger training set will help.



(a) (b)

Figure 4: (a) 2D model slice on which we test our classification algorithm. (b) Classification results. While we seem to retrieve the approximate cavern location, the classifier returns many false positives. [**CR**] fantine1/. test,result

DISCUSSION

The classification results indicate that it is possible to retrieve approximate karst cavern locations from seismic recordings using pattern recognition algorithms. Migration is an essential step for accurate labeling of caverns. However, the classification tests returned many false positives. These results seem to indicate that the classifier would benefit from using more input features and larger training data. Another idea would be to change classification algorithm for an edge detection method.

The next step would be to explore whether a classifier could be trained on synthetic data, which can be conveniently labeled, and be used to predict cavern locations on field data. We were provided a with a three-dimensional mapping of the cavern locations of the Lechuguilla channel system from the Carlsbad Caverns National Park. It features a fine and intricate channel system with caverns of various scales. This would allow us to generate data on a configuration that is closer to reality than our simplistic synthetic model.

ACKNOWLEDGMENTS

We would like to thank Gustavo Catao Alves for his help with his elastic modeling code.

REFERENCES

Alves, G., 2015, Adjoint formulation for the elastic wave equation: SEP-Report, **158**, 133–150.

Berkovitch, A., I. Belfer, Y. Hassin, and E. Landa, 2009, Diffraction imaging by multifocusing: Geophysics, **74**, WCA75–WCA81.

Biondi, B., 2006, 3d seismic imaging: Society of Exploration Geophysicists.

Collobert, R., S. Bengio, and J. Mariéthoz, 2002, Torch: a modular machine learning software library: Technical report, IDIAP.

Fisher, R. A., 1936, The use of multiple measurements in taxonomic problems: Annals of eugenics, **7**, 179–188.

Fomel, S., E. Landa, and M. T. Taner, 2007, Poststack velocity analysis by separation and imaging of seismic diffractions: Geophysics, **72**, U89–U94.

Hastie, T., R. Tibshirani, J. Friedman, and J. Franklin, 2005, The elements of statistical learning: Data mining, inference and prediction: 534–552.

Hsu, C.-W., C.-C. Chang, C.-J. Lin, et al., 2003, A practical guide to support vector classification.

Huot, F., 2016, Exploratory data analysis of the sola land dataset: SEP–Report, **165**.

Huot, F. and R. Clapp, 2016, Detecting karst caverns by pattern recognition : SEP-Report, **163**, 297–307.

James, G., D. Witten, T. Hastie, and R. Tibshirani, 2013, An introduction to statistical learning: Springer.

Jia, Y., E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, 2014, Caffe: Convolutional architecture for fast feature embedding: Proceedings of the ACM International Conference on Multimedia, 675–678.

Lester, R., E. Liebes, R. Hill, S. Jenkins, A. Makedonov, et al., 2015, Investigating mega-amplitudes in tengiz carbonates through interactive model-building and gaussian beam migration.

Osuna, E., R. Freund, and F. Girosi, 1997, Training support vector machines: an application to face detection: Computer vision and pattern recognition, 130–136.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., 2011, Scikit-learn: Machine learning in python: The Journal of Machine Learning Research, **12**, 2825–2830.

Stanford Exploration Project, SEP165, October 8, 2016

Automatic wave mode identification using machine learning

Ohad Barak and Fantine Huot

ABSTRACT

In addition to reflection data, seismic recordings contain many different wave modes that are either unwanted or unneeded, degrading overall data quality. We use support vector machines (SVM), a type of supervised learning algorithm, for automatic wave mode classification. We decompose multicomponent translational and rotational seismic data from a field survey into polarization vectors, by applying continuous wavelet transforms (CWT) followed by singular value decomposition (SVD). We train an SVM classifier to distinguish surface waves from body waves from these polarization vectors, and show classification results on different portions of the field data. Our method does not rely on spatial continuity, and can therefore be applied to spatially aliased data.

INTRODUCTION

In addition to reflection data, seismic surveys are cluttered with many other seismic responses that are either unwanted or unneeded, resulting in recorded data containing many different wave modes (Yilmaz, 2001). In the case of land acquisitions, high amplitude ground roll noise can obscure signal, degrading overall data quality.

Ground roll is the main type of coherent noise in land seismic surveys and is characterized by low frequencies and high amplitudes. Common processing techniques for attenuating ground roll include frequency filtering (Yilmaz, 2001), Radon transform (Liu and Marfurt, 2004), wavelet transforms (Deighan and Watts, 1997), and the curvelet transform (Yarham and Herrmann, 2008). However, these techniques can be limited when ground roll is spatially aliased and has non-linear moveout, and it is notoriously difficult to model ground roll with sufficient generality. Therefore, ground noise removal remains a tedious task.

Barak and Ronen (2016) demonstrate how to use combined translational and rotational data to identify and separate particular wave modes. They apply continuous wavelet transforms (CWT) followed by singular value decomposition (SVD) to identify the polarization signature of the particular wave modes associated with ground roll. A filter is then applied to attenuate sections of the data with similar polarization signatures.

Nonetheless, this methodology requires selective manual picking of events on the data section. Moreover, it is based on the underlying assumption of stationarity of wave modes along offset, although there is no guarantee that the polarization signature of a certain wave mode at a particular offset should remain similar at other offsets. Therefore, we extend this process by incorporating pattern recognition algorithms as described by Huot and Clapp (2016). We use support vector machines (SVM), a set of supervised learning methods used for classification, to identify the waves modes associated with ground roll, at all times, offsets and azimuths.

We train an SVM classifier to identify specific wave modes from the polarization vectors computed on field data, and then test our classification on different portions of the data where the wave modes are clearly distinguishable to assess the quality of our classification.

PREPARING DATA FEATURES FOR MACHINE LEARNING

Continuous Wavelet Transform (CWT)

The continuous wavelet transform (CWT) is defined as:

$$C_i\left(a,b;g_i(t),\psi(t)\right)=\int_{-\infty}^{\infty}g_i(t)\frac{1}{\sqrt{a}}\psi^*\left(\frac{t-b}{a}\right)dt, \tag{1}$$

where $g_i(t)$ is the input signal of the i^{th} data component, $\psi(t)$ is a mother wavelet, ψ^* is a daughter wavelet, which is the complex conjugate of the mother wavelet stretched by scale a and time-shifted by b. It is common to use the Morlet wavelet as a mother wavelet, and we do so in this paper. For brevity, we will use $C_i\left(a,b\right) := C_i\left(a,b;g_i(t),\psi(t)\right)$

The continuous wavelet transform effectively shows how correlated our time-series is with a particular daughter wavelet. Since the correlation is done in running time windows (shifted by b), the transform retains the temporal sense of the data and yet decomposes it to wavelet scales, which are in essence similar to frequency. We use this time-frequency decomposition to identify wave modes of particular frequencies that appear at particular times in multicomponent data.

Singular Value Decomposition (SVD)

We apply singular value decomposition to a time slice of a single continuous-wavelet transformed multicomponent trace $C_i(a,b_k)$, where b_k is the time index of the slice, a ranges through the wavelet scales and i represents the data component. Therefore, we have an $N_a \times N_c$ data matrix $\mathbf{D}$ where the rows are the wavelet scales and the columns are the data components. SVD is a method of finding the waveform $\mathbf{u}$, magnitude σ, and polarization $\mathbf{v}$ of the signal that is present in the data $\mathbf{D}$. The SVD of the data $\mathbf{D}$ is given by

$$\mathbf{D}=\mathbf{U}\Sigma\mathbf{V}^T, \tag{2}$$

where $\mathbf{D}$ is the product of an $N_a \times N_c$ matrix $\mathbf{U}$, an $N_c \times N_c$ diagonal matrix Σ, and the transpose of an $N_c \times N_c$ matrix $\mathbf{V}$. The unit left and right singular vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ are the column-vectors of $\mathbf{U}$ and $\mathbf{V}$. The singular values σ_i are the diagonal elements of Σ. They are ordered such that $|\sigma_1|$ is the greatest and $|\sigma_{N_c}|$ the smallest. The left and right singular vectors are mutually orthogonal, such that $\mathbf{U}^T\mathbf{U}=\mathbf{I}$ and $\mathbf{V}\mathbf{V}^T=\mathbf{I}$.

The right singular vectors $\mathbf{v}_i$ display the polarization of the data within the particular frequency window along the data axes. We transpose and multiply the matrix $\mathbf{V}$ by the singular value matrix Σ, to obtain the scaled polarization vectors:

$$\mathbf{S}=\boldsymbol{\Sigma}\mathbf{V}^T. \tag{3}$$

Support Vector Machines (SVM)

The polarization vectors obtained after applying CWT and SVD are used as input for classifying the different wave modes. From here, we set up the classification problem according to the methodology described in Huot and Clapp (2016).

Support vector machines (SVM) are a supervised learning algorithm that have been shown to perform well in a variety of settings, and are often considered one of the best "out of the box" classifiers. To train the classifier, we build a training dataset of N sample pairs, $(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)$, where the $x_i \in \mathbb{R}_p$ represent the features, which are the p values of the polarization vectors, and the $y_i \in \{-1, 1\}$ are the binary labels indicating to which wave mode the sample belongs, namely, whether it is a surface wave or a body wave.

The support vector classifier determines the optimal hyperplane separating the two classes in the features space, given a certain amount of slack $\xi = (\xi_1, \xi_2, \cdots, \xi_N)$ to account for the fact that the samples are not perfectly separable. If we define a hyperplane by the following:

$$\{x : f(x) = x^T\beta + \beta_0 = 0\}, \tag{4}$$

where β is a unit vector and β_0 a constant, the classification rule induced by $f(x)$ can be expressed as:

$$G(x) = \text{sign}[x^T\beta + \beta_0]. \tag{5}$$

With these notations, the support vector classifier for the nonseparable case is commonly expressed as:

$$\min_{\beta,\beta_0} \|\beta\| \quad \text{subject to} \left\{ \begin{array}{l} y_i({x_i}^T\beta + \beta_0) \geq 1 - \xi_i\,,\ \forall i \\ \xi_i \geq 0,\ \forall i \quad \text{and} \quad \sum \xi_i \leq K. \end{array} \right. \tag{6}$$

By bounding the sum $\sum \xi_i$, we bound the total proportional amount by which predictions fall on the wrong side of their margin. Misclassifications occur when $\xi_i > 1$, so bounding $\sum \xi_i$ at a value K bounds the total number of training misclassifications at K.

However, the support vector classifier subsequently described is limited to finding linear boundaries in the input feature space. To achieve better training-class separation, we introduce the idea of enlarging the feature space using basis expansions. Linear boundaries in the enlarged space translate to nonlinear boundaries in the original space, thereby making the problem more flexible. Once the basis functions $h_m(x)$, $m = 1, \cdots, M$ are selected, the procedure is the same as before. We fit the support vector classifier using input features $h(x_i) = (h_1(x_i), h_2(x_i), \cdots, h_M(x_i))$, $i = 1, \cdots, N$, and produce the (nonlinear) function $\hat{f}(x) = h(x)^T\hat{\beta} + \hat{\beta}_0$. The classifier is $\hat{G}(x) = \text{sign}[\hat{f}(x)]$ as before.

In practice, we need not specify the transformation $h(x)$ at all, but require only knowledge of the following kernel function:

$$K(x, x') = \langle h(x), h(x') \rangle, \tag{7}$$

that computes the inner products in the transformed space. For particular choices of h, these inner products can be computed very efficiently.

Three popular choices for K in the SVM literature are shown in the following:

$$
\begin{aligned}
d^{\text{th}}\text{—degree polynomial:} \quad K(x,x') &= (1+\langle x,x'\rangle)^d, && (8)\\
\text{Radial basis:} \quad K(x,x') &= \exp(-\gamma\|x-x'\|^2), && (9)\\
\text{Neural network:} \quad K(x,x') &= \tanh(\kappa_1\langle x,x'\rangle+\kappa_2). && (10)
\end{aligned}
$$

A more complete overview of the implementation of support vector machines is described in Huot and Clapp (2016).

SVM TRAINING AND TESTING WITH KETTLEMAN DATA

The 2D Kettleman survey comprised multiple types of sources and multiple types of receivers, both on the surface and at depth. The shot line length was 1.6 km long. The seismic sources used were a vibroseis truck, an accelerated weight-drop, and buried dynamite charges. At one end of the shot line there were five 3 component (3C) linear accelerometers, which were closely spaced at a 2.1 m interval inline. Additionally, near the center of the shot line there were two adjacent 3C geophones buried at 1 m depth, spaced at a 2 m interval inline.

Although the survey also had 3C rotation sensors, the signal to noise ratio of these sensors was very low for offsets greater than 300 m. Therefore, we derived the pitch rotational component by differencing adjacent vertical geophones and accelerometers at each receiver station. Derivation of rotational components from translational data is described in Barak and Ronen (2016), Barak et al. (2015), Muyzert et al. (2012) and Brokesova and Malek (2015).

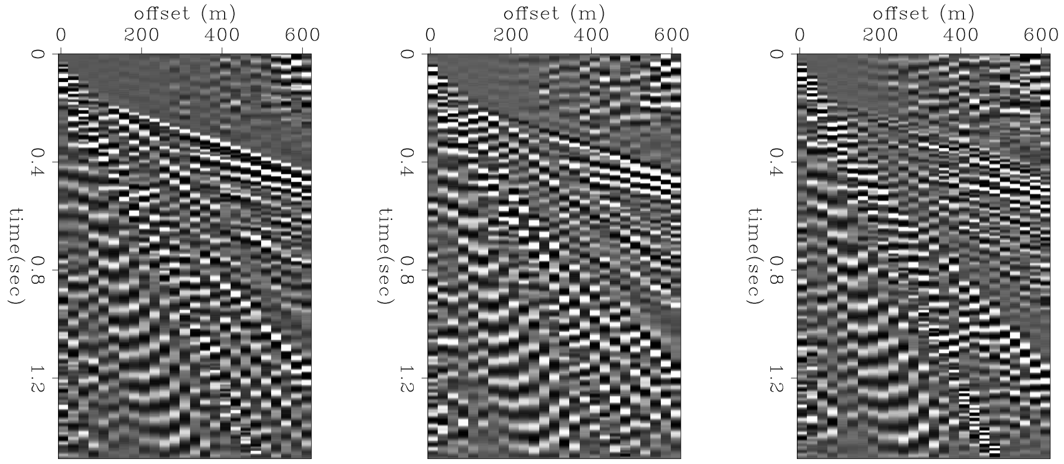
Since the sensors were arrayed only in the inline direction, we could not derive rotational data for the roll and yaw components. Consequently, our analysis here includes only the vertical and radial translational components and the pitch rotational component. Figure 1(a) is the vertical component receiver gather at station 335 shot with the vibroseis source, Figure 1(b) is the radial component, and Figure 1(c) is the pitch component. AGC has been applied for display.

The P body-wave reflections are visible at earlier times and longer offsets, but most of the section is dominated by various modes of surface waves, which is typical in land data. There is a very slow Rayleigh wave mode propagating at around 250 m/s, and a faster mode propagating at 400 m/s. Henceforth we will refer to these wave modes as “slow” and “fast” ground roll. There is yet another, faster mode, propagating at 600 m/s - 700 m/s, which is not part of the analysis here.

Standard seismic processing requires that the surface waves (“ground-roll”) be removed, as the data we are commonly interested in are the P body-wave reflections that are obscured by the surface waves. However, as a first step, the unwanted energy must be identified.

SVM training data for classification of surface waves

In order to separate the surface waves from the body waves, we train an SVM classifier on the data from receiver station 335. We decompose the training data into scaled continuous-wavelet polarization vectors by applying CWT followed by SVD, as in equations 1 and 3.



(a) (b) (c)

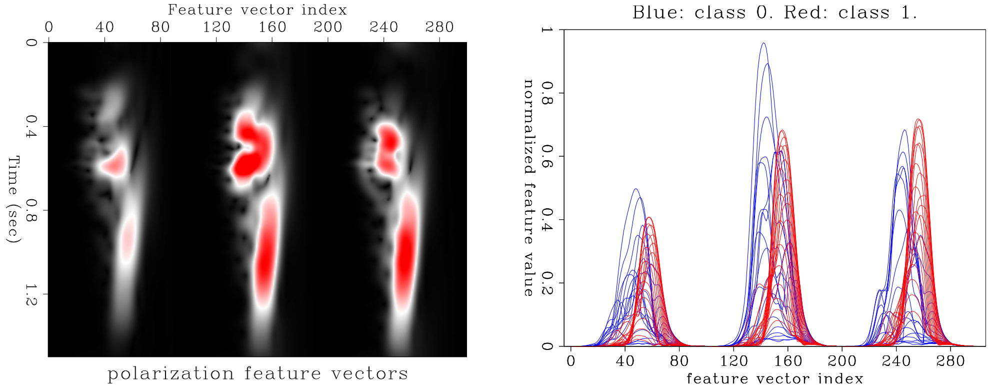
Figure 1: Receiver gathers at station 335 for the vibroseis source. a) Vertical accelerometer component. b) Radial accelerometer component. c) Pitch component derived from differencing two adjacent inline vertical accelerometers. AGC has been applied for display. Note the various types of surface waves present. [**ER**] ohad1/. Vib65-st335-vz-bp,Vib65-st335-vx-bp,Vib65-st335-vz-dx

Since we are using three data components, for each offset and each time sample we obtain three polarization vectors. We use only the first polarization vector as our input feature vector for the SVM classification, since most of the energy is contained within the first vector.

The feature vectors are illustrated in Figure 2(a). We can clearly distinguish high amplitude parcels corresponding to the specific seismic signature of the different wave modes. Figure 2(b) represents the same feature vectors plotted by color code, red for slow ground roll and blue for all other wave modes. We can see that the wave modes have distinct behaviors in the feature space.

In order to train the classifier, we label the training data manually by selecting particular wave modes in the data from receiver station 335. The selection was done using a linear mute function that windowed only the ground-roll energy. Figures 3(a), 3(b) and 3(c) illustrate which regions were labeled as slow ground roll, fast ground roll and total ground roll. For each of these selected wave modes, we trained an SVM classifier using the aforementioned polarization feature vectors as input. The ground roll is labeled as “Class 1”, while all the other wave modes are labeled as “Class 0”.

We then test the SVM’s fitting capability. The classification prediction results obtained on the training data are illustrated in Figures 3(d), 3(e) and 3(f). They are highly similar to our hand-picked labels, indicating that our classifiers fit the training data well, with a low training error.



(a) (b)

Figure 2: The training data at receiver station 335 for the SVM. a) Feature vectors for each time sample at offset 180m. b) The same feature vectors color coded by their respective classification, 'Class 1' being the slow ground roll mode and 'Class 0' representing everything else. [**ER**] ohad1/. Vib65-st335-3cr-features-V1-off180a,Vib65-st335-3cr-features-V1-off180b

SVM prediction of surface waves on test data

Testing on other receiver stations

We then tested our previously trained classifiers on data from other receiver stations for the same vibroseis source type.

Stations 336 and 337 are adjacent to stations 335 (the training data station) near the end of the shot line, and therefore the data at these receiver stations should be similar to the training data. This enables us to estimate how well the classifier predicts the ground roll wave mode on these test data. Figure 4(a) is the vertical component at station 336. Figures 4(b), 4(c) and 4(d) are the predicted slow ground roll, fast ground roll and combined slow+fast ground roll wave modes at this station, respectively. Recall that this prediction is done using the classification generated by labeling the data at station 335. Observe that the classifier has managed to identify the ground roll in these test data relatively well, and it does not misclassify the body wave energy as ground roll.

The prediction seems to be better when both ground roll modes are combined, as show in Figure 4(d). It seems that it is more difficult for the classifier to differentiate between the two types of ground roll (slow vs fast) than between the ground roll and body waves. The same conclusions can be drawn for the prediction results for station 337, shown in Figures 4(f), 4(g) and 4(h).

Station 191 is about 1 km away from station 335. At station 191 there were geophones installed, rather than accelerometers as at stations 335, 336 and 337. Furthermore, the geophones were buried at 1 meter depth, whereas at stations 335, 336 and 337 the accelerometers were on the surface.

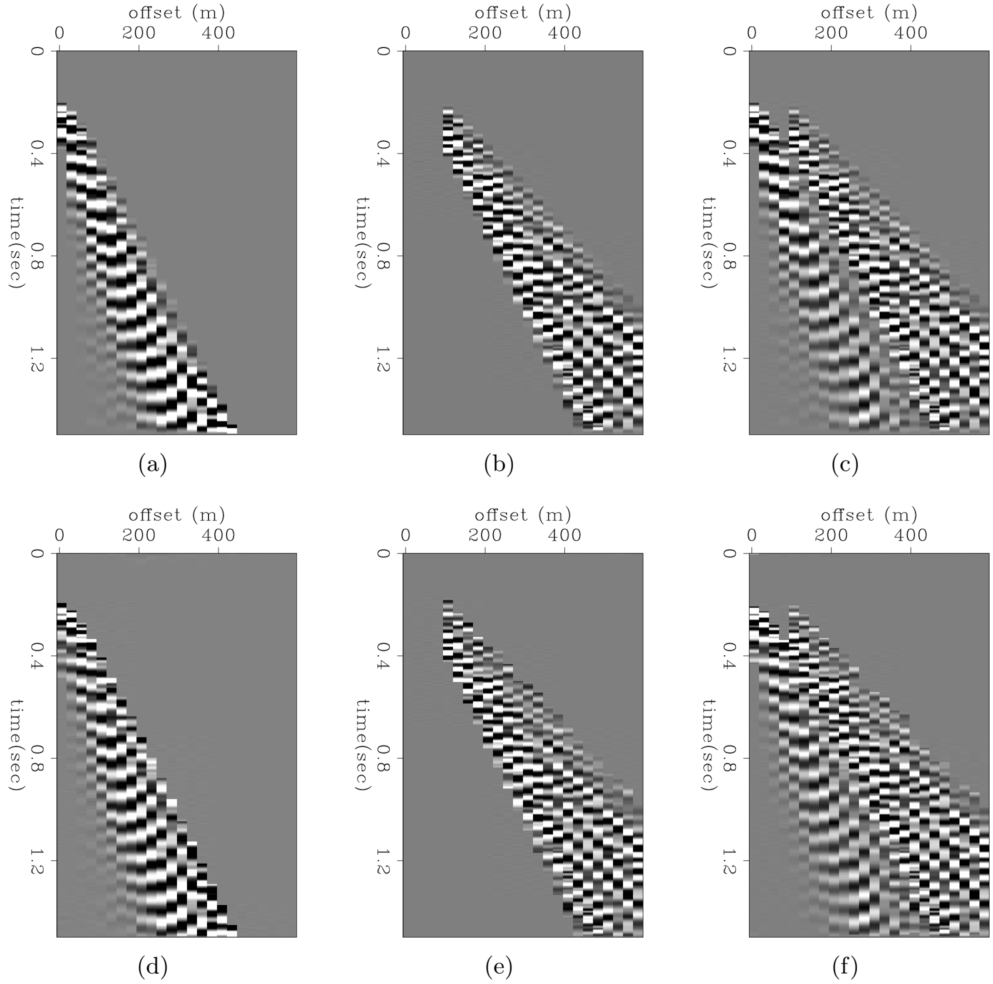


Figure 3: SVM training data and fitting results from station 335, for the vibroseis source. a) Slow ground roll mode training data. b) Fast ground roll mode training data. c) Combined (slow+fast) ground roll modes training data. d) SVM fitting of slow ground roll mode on training data. e) SVM fitting of fast ground roll mode on training data. f) SVM fitting of combined (slow+fast) ground roll modes on training data. [**ER**] ohad1/. 3a,3b,3c,3d,3e,3f

Figures 4(j), 4(k) and 4(l) are the predicted slow ground roll, fast ground roll and combined slow+fast ground roll wave modes at station 191, respectively. The classifier seems to have more difficulty differentiating between the slow and the fast ground roll modes at this station. Additionally, it has misclassified a small portion of the body wave energy as ground roll, as can be seen on the left Figure 4(l) at $t = 0.4$ s. However, by and large, the classification does enable the identification of ground roll energy, despite the differences between the data at station 191 vs station 335, on which we trained our SVM classifier.

Testing on other source types

We then tested the classifier on the same station that provided the training data (335), but for alternate seismic source types.

Figures 5(a) and 5(d) are again the receiver gather of the vertical component at station 335, and portion of combined (slow+fast) ground roll labeled as "Class" 1 for the training data.

Figure 5(b) is the receiver gather of the vertical component at station 335 for an accelerated weight-drop source. Note that the sampling for this source type was much denser than the vibroseis source. The ground roll wave modes visible in this section are similar to those in Figure 5(a), however they are not aliased. Figure 5(e) shows the SVM's prediction of the combined (slow+fast) ground roll modes for the accelerated weight drop data. Despite the fact that the source type is different, the classification is able to correctly predict the ground roll in this section. Figure 5(e) is remarkably similar to Figure 5(d), albeit with a better sampling.

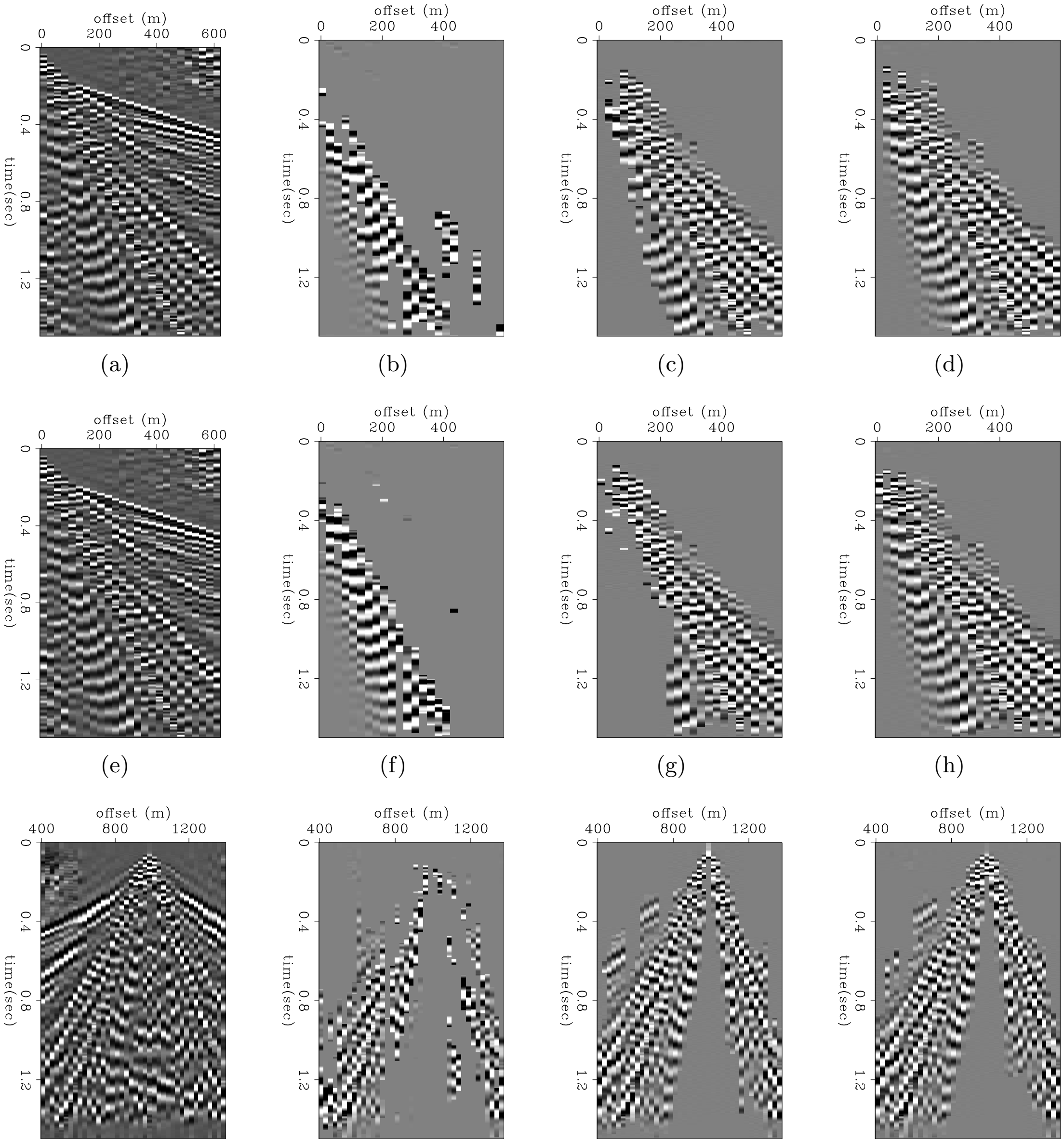
Both the vibroseis and the accelerated weight-drop source were on the surface. Figure 5(c) is the receiver gather at station 335, but for a dynamite source buried at 25 m depth. The ground roll energy has a different appearance in this section. Figure 5(f) is the SVM prediction of the slow+fast ground roll wave mode for the dynamite source. Note that the energy classified as ground roll does indeed seem to have the same linear moveout as the ground roll in the vibroseis section. The classifier even picks up on a portion of a faster surface wave mode that does not appear in the vibroseis data.

In both test cases shown in Figures 5(e) and 5(f), there is no misclassification of the body wave energy as ground roll.

DISCUSSION

Standard seismic processing may utilize some known attributes of ground roll to identify it. For example, ground roll has a much slower moveout than body waves in common shot or receiver gathers. Additionally, in multicomponent data the elliptical polarization associated with Rayleigh waves may be employed to identify ground roll.

The approach of machine learning is very different however. Standard methods use an a priori, analytical model of ground roll, such as moveout or elliptical polarization. A

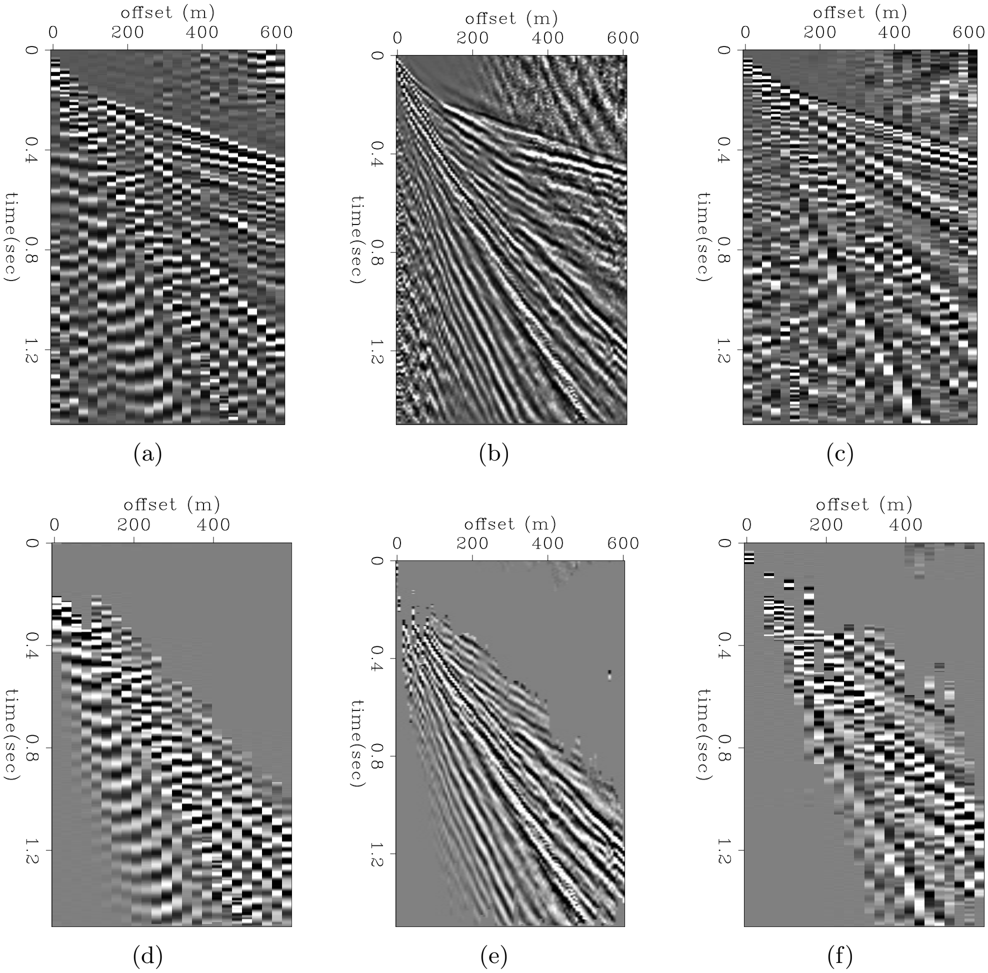


(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure 4: SVM prediction of surface wave modes on test data, for the vibroseis source. a,e,i): Vertical component at receiver stations 336, 337 and 191. b,f,j): prediction of slow ground roll mode for stations 336, 337 and 191. c,g,k): prediction of fast ground roll mode for stations 336, 337 and 191. d,h,l): prediction of combined (slow+fast) ground roll modes for stations 336, 337 and 191. [**ER**] ohad1/. 4a,4b,4c,4d,4e,4f,4g,4h,4i,4j,4k,4l



(a) (b) (c)

(d) (e) (f)

Figure 5: SVM prediction of surface wave modes at station 335, for different seismic source types. a) Vertical component of vibroseis source. b) Vertical component of accelerated weight-drop source. c) Vertical component of dynamite source. d) The portion of the combined (slow+fast) ground roll mode labeled as class 1 from the vibroseis source data. e) prediction of combined (slow+fast) ground roll mode for the accelerated weight-drop source. f) prediction of combined (slow+fast) ground roll mode for the dynamite source. [**ER**] ohad1/. 5a,5b,5c,5d,5e,5f

machine learning approach does away with any predetermined model, and substitutes a model learned from the data themselves. The advantage of such an approach is that it may work in cases where a useful, representative model is beyond our capability to represent analytically.

The case of surface waves (ground roll) is of particular interest in this respect, since the near surface of the Earth is commonly very complex, and in general it is difficult to accurately model wave propagation in the near surface. For example, Rayleigh waves do not always have elliptical polarization, nor are their moveouts necessarily predictable. Surface waves may also be generated by body waves incident on near surface scatterers, in addition to being radiated directly away from the source position. Therefore, the prospect of enabling an algorithm to learn from the data what ground roll may look like in multiple scenarios is compelling.

We have used the continuous-wavelet polarization vectors of multicomponent seismic data to train an SVM algorithm to identify ground roll. The classification results using the test data indicate that the SVM is indeed able to identify ground roll based on this (rather minimal) training.

There remain several open questions with respect to practical application of this machine learning algorithm:

1. In order to have a good representation of the various ways each wave mode may be polarized at multiple times/offsets/azimuths (given a set of data components), massive amounts of seismic data are required. How can we classify massive quantities of data?
2. Can an SVM trained on a particular dataset be used for classification of another dataset?
3. Are more components necessarily better for the SVM's classification of wave modes? If not, which minimal set of components would be the most useful?

The Kettleman dataset is instrumental in showing the possibility of identifying wave modes in land data using machine learning, since it has multiple components, both translational and rotational, which enable identification by polarization. However, the Kettleman dataset is very small, and therefore we cannot answer the questions stated above with it. For that, we would require a very large, 3D multicomponent land or OBS dataset.

ACKNOWLEDGEMENTS

We thank Chevron for permission to publish the Kettleman dataset.

REFERENCES

Barak, O., K. Key, S. Constable, P. Milligan, and S. Ronen, 2015, Acquiring rotation data on the ocean bottom without rotation sensors: SEP-Report, **158**, 311–322.

Barak, O. and S. Ronen, 2016, Wave-mode separation in the complex wavelet domain using combined translational and rotational data : SEP-Report, **163**, 1–19.

Brokesova, J. and J. Malek, 2015, Six-degree-of-freedom near-source seismic motions II: Examples of real seismogram analysis and S-wave velocity retrieval: Journal of Seismology, **19**, 511–539.

Deighan, A. J. and D. R. Watts, 1997, Ground-roll suppression using the wavelet transform: Geophysics, **62**, 1896–1903.

Huot, F. and R. Clapp, 2016, Detecting karst caverns by pattern recognition : SEP-Report, **163**, 297–307.

Liu, J. and K. J. Marfurt, 2004, 3D high resolution Radon transforms applied to ground roll suppression in orthogonal seismic surveys: SEG Technical Program Expanded Abstracts, 2144–2147.

Muyzert, E., A. Kashubin, E. Kragh, and P. Edme, 2012, Land seismic data acquisition using rotation sensors: 74th Conference and Exhibition, EAGE, Extended Abstracts.

Yarham, C. and F. J. Herrmann, 2008, Bayesian ground-roll separation by curvelet-domain sparsity promotion: SEG Technical Program Expanded Abstracts, 2576–2580.

Yilmaz, O., 2001, Seismic data analysis: Society of Exploration Geophysicists, Tulsa, **1**.

Detection and removal of coherent anthropogenic noise from passive seismic data

Ethan Williams and Eileen Martin

ABSTRACT

We analyze the impact of identifying and removing coherent anthropogenic noise on synthetic Green's functions extracted from ambient noise recorded on a shallow trenched, dense, linear distributed acoustic sensing (DAS) array. Low-cost, low-impact urban seismic surveys are possible with ambient noise recorded by DAS, which uses dynamic strain sensing to detect seismic waves incident to a buried fiber optic cable. However, ambient noise data recorded in urban areas include coherent, time-correlated noise from near-field infrastructure such as cars and trains passing the array, in some cases causing artifacts in estimated Green's functions and yielding potentially incorrect surface wave velocities. Based on our comparison of several methods, we propose an automated, real-time data processing workflow to detect and reduce the impact of these events on data from a dense array in an urban environment. We show the effect of removing such unwanted noise on estimated Green's functions from ambient noise data recorded in Richmond, CA in December 2014 and Fairbanks, AK in August 2015.

INTRODUCTION

Urban seismic surveys are an essential tool for many areas of geoscience and civil engineering, including the design of earthquake-resistant structures and the quantification of seismic hazard in cities. However, conventional seismic surveys are all but impossible in the urban environment because of the high impact of active sources and the difficulty of deploying a sizable receiver array. Recent experiments have successfully shown the applicability of ambient noise interferometry on recordings made with distributed acoustic sensing (DAS) as an alternative to conventional surveys that is both low-cost and low-impact (Martin et al., 2015; Daley et al., 2013; Ajo-Franklin et al., 2015). In a DAS survey, an interrogator unit regularly transmits a coherent laser pulse along a fiber optic cable network and records the Rayleigh backscatter intensity, which is converted into an approximate seismogram. The ambient noise recorded using DAS can then be correlated to extract synthetic Green's functions through seismic interferometry, as has been successfully done by Martin et al. (2015) and Ajo-Franklin et al. (2015).

One important barrier to the wide application of DAS surveys in urban areas is the presence of coherent transportation-related noise from sources such as cars or trucks passing near the array. During correlation, these events correlate with themselves to produce artifacts in Green's functions (Martin et al., 2015). In order to improve Green's function estimates in these environments, there have been efforts to identify and automatically down-weight high amplitude noise (Lindsey, 2016). Here, we build on these methods more specifically to identify strong moving sources of noise related to transportation. First, we propose a real-time processing workflow that develops a catalog of times where strong transportation-related noise has been recorded. Our method utilizes the short-term average over long-term average (STA/LTA) algorithm from the field of earthquake detection and takes advantage of the density of the data recordings to minimize the number of false triggers. Second, we demonstrate a method for down-weighting these events during interferometry that improves the strength and coherency of Green's functions estimates on the test data. The methods discussed in this paper have been developed for use with real-time traffic monitoring and the removal of near-array transportation-related noise in the DAS ambient noise survey on Stanford campus that began in September 2016.

CHARACTERIZING NOISE SOURCES

Several sources of noise contribute to any ambient noise recording. Generally, incoherent noise from microseism dominates at low frequencies and incoherent anthropogenic noise from freeways and other infrastructure dominates at high frequencies. Additionally, recordings often include signals from earthquakes and coherent noise from infrastructure near the array, such as passing cars.

Seismic interferometry posits that the correlation of a diffuse wavefield in a medium will return the impulse response of that medium (Wapenaar et al., 2010). Applied to the Earth, this means that the cross-correlation of ambient noise recorded on different channels in an array with that recorded on a single channel produces an approximate Green's function, which represents the response of the Earth to a receiver acting as a virtual source. Greens functions constructed from ambient noise can be used to monitor changing properties in the subsurface or for tomographic investigations. The presence of coherent, time-correlated noise in an ambient noise recording violates the diffuse wavefield assumption and results in artifacts in the estimated source response that resemble multiple virtual sources.

Ambient noise recorded using DAS in Richmond, CA in 2014 and in Fairbanks, AK in 2015 contained visually identifiable coherent noise from passing traffic (Ajo-Franklin et al., 2015; Martin et al., 2016). Incoherent noise from urban infrastructure far from the recording array is the primary noise source used for high frequency ambient noise interferometry because it is a diffuse component of the ambient spectrum (e.g. Nakata et al. (2011)), but coherent noise from traffic sources near the recording array is the primary cause of artifacts. This is because traffic passing the array is in two ways time-correlated. First, surface waves generated by a passing vehicle have visually identifiable group moveout across the array at traffic speed, giving the wave package associated with any given car a temporally-correlated velocity relationship. Second, surface waves generated by the vehicle do not originate at the source channel but rather at road bumps and other features that lie along the array, causing additional virtual sources within the response estimate. An example recording of a vehicle passing the Fairbanks array is shown in Figure 1.

The geometry of relevant roads and their distance from the array determine the properties of the coherent transportation-related noise recorded. In Fairbanks, the closest cars travelled on a road very near and parallel to the array, so that their traffic speed was easily identifiable as the group move out of the ground roll at approximately 25 m/s. In Richmond, the closest cars travelled on a road slightly farther from and oblique to the array, so that the traffic speed was difficult to discern but the surface waves were still highly coherent.

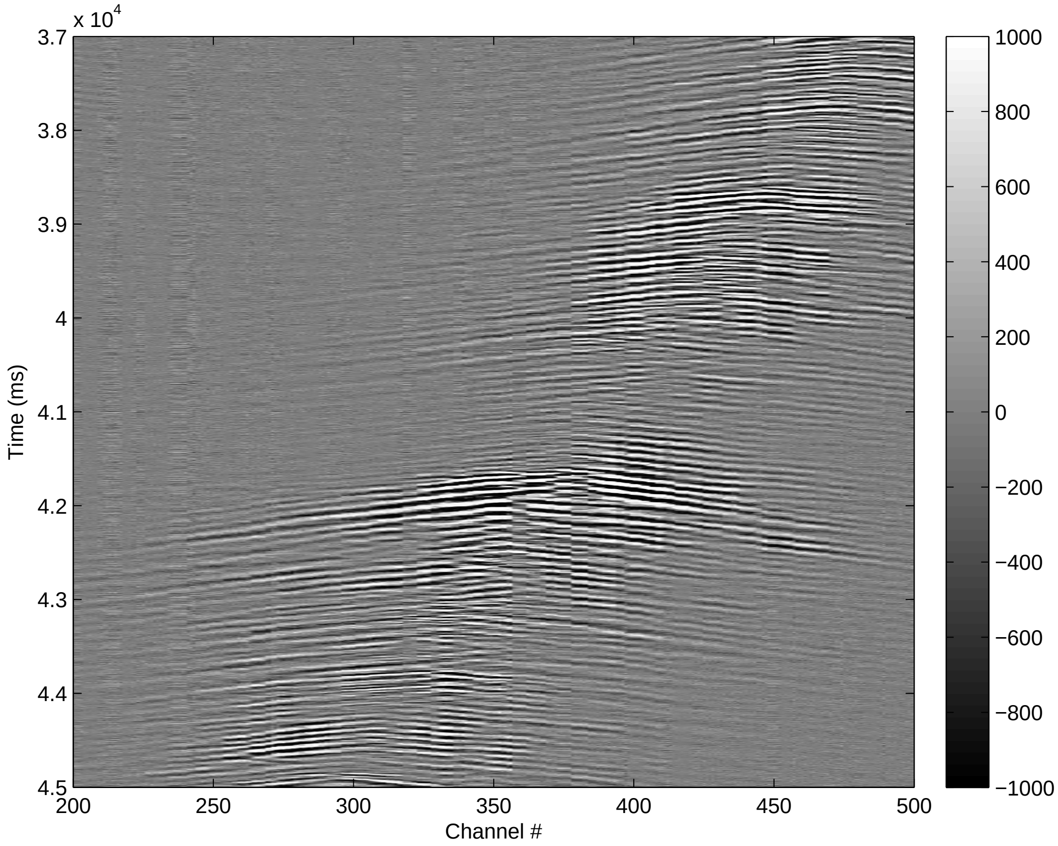


Figure 1: Example recording showing a car driving past a linear segment of the trenched DAS array in Fairbanks, AK. Channel spacing is 1 m. [**CR**] ethan1/. example-event

The following methods were designed on both data sets, and we present suggestions for modifying proposed processing workflow to best suit recordings with different geometries and noise intensities.

COMPARISON OF METHODS

In the earthquake detection literature, three principal methods for event detection exist: (1) autocorrelation, (2) waveform comparison methods, and (3) short-term average over long-term average (STA/LTA) (Withers et al., 1998).

One prominent feature of data recorded with DAS is optical noise spikes at random intervals over several channels due to vibration of the laser interrogator unit. The autocorrelation method for event detection entails partitioning each channel's record into short-time windows and cross-correlating all possible combinations of these windows. For data recorded with DAS, optical noise correlates with itself, creating artificially high correlation coefficients and making transportation-related events difficult to distinguish. Even more importantly, this method can only operate on a single channel, so performing this method across an array of several thousand channels proves to be too slow to run in real time with ordinary computational resources.

While waveform comparison methods have been shown to be the quickest and most effective method of identifying earthquakes (Yoon et al., 2015), waveform comparison tests are impossible to implement for data recorded with DAS because there are no catalogs of transportation-related event recordings to build a database. Because of DAS's low signal-to-noise ratio compared to conventional seismic recording instruments and because the traces themselves are only derived indirectly from strain rate approximations, ambient noise recorded with DAS cannot be compared with conventional data.

The STA/LTA method effectively detects impulsive events with high signal-to-noise ratio by comparing the squared amplitudes of data summed over two moving windows of differing lengths. Equation 1 takes two parameters—STA window length (ℓ_{sta}) and LTA window length (ℓ_{lta})—which need to be set manually based on the signal-to-noise ratio and other parameters of the data. The long-term window finds the average amplitude of the data for some number of samples ℓ_{lta} before sample i, which is representative of the background, and the short-term window finds the average amplitude of the data for a shorter number of samples ℓ_{sta} before sample i, which is representative of the signal at i. When amplitude increases significantly at some iteration of sample i, the STA/LTA ratio increases. Thus, in conventional earthquake detection, STA/LTA is a trigger-based algorithm, meaning that the user pre-sets a threshold STA/LTA value and when the STA/LTA ratio calculated on a seismogram exceeds this value, an event is cataloged. The STA/LTA method is advantageous for real-time event detection simultaneously across channels on a large recording array because it can be easily vectorized across channels and it is often implemented using a recursive algorithm, making it the cheapest detection method by far.

$$\frac{STA}{LTA}(i) = \frac{\sum\limits_{j=i-\ell_{sta}}^{i} |u(j)|^2}{\sum\limits_{j=i-\ell_{lta}}^{i} |u(j)|^2} \tag{1}$$

PROCESSING WORKFLOW

STA/LTA

We first import a single 60 s file and apply basic pre-processing steps, including a despiking tool to reduce the number of optical noise spikes. Because our data were recorded at $f_s = 1$ kHz, we lowpass filter the data and downsample by a factor of 10, which improves computational speed without reducing the quality of the STA/LTA image. We then calculate STA/LTA on all channels of the file for the complete record length using a recursive algorithm (Equation 2) that approximates the actual STA/LTA ratio very closely.

$$\begin{aligned} STA(i) &= \frac{|u(j)|^2}{\ell_{sta}} + (1 - \frac{1}{\ell_{sta}}) \times STA(i-1) \\ LTA(i) &= \frac{|u(j)|^2}{\ell_{lta}} + (1 - \frac{1}{\ell_{lta}}) \times LTA(i-1) \end{aligned} \tag{2}$$

We found that $\ell_{sta} = 0.5$ s and $\ell_{lta} = 10$ s were effective parameter choices for data recorded in both Richmond and Fairbanks. Smaller ℓ_{sta}/ℓ_{lta} reduces the event-to-background contrast, whereas higher ℓ_{sta}/ℓ_{lta} reduces the sharpness of event arrivals (see Figure 2).

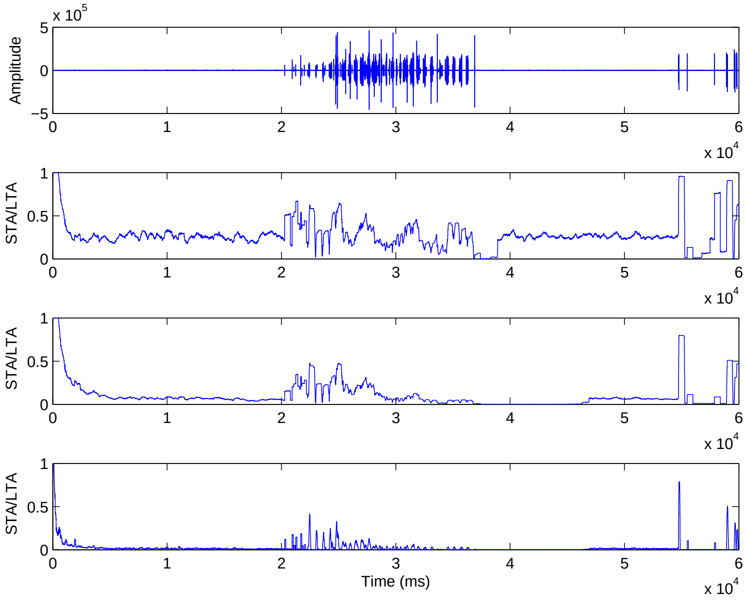


Figure 2: Examples of different ℓ_{sta} and ℓ_{lta} parameter choices. Data must be padded with $\ell_{lta} - 1$ samples to prevent the spike at the beginning, which occurs because the LTA window is only partially filled. From the top, (a) the original trace (raw data), (b) $\ell_{sta} = 0.5$ s, $\ell_{lta} = 2$ s, (c) $\ell_{sta} = 0.5$ s, $\ell_{lta} = 10$ s, (d) $\ell_{sta} = 0.1$ s, $\ell_{lta} = 10$ s. **[CR]** ethan1/. stalta-examples

Window check

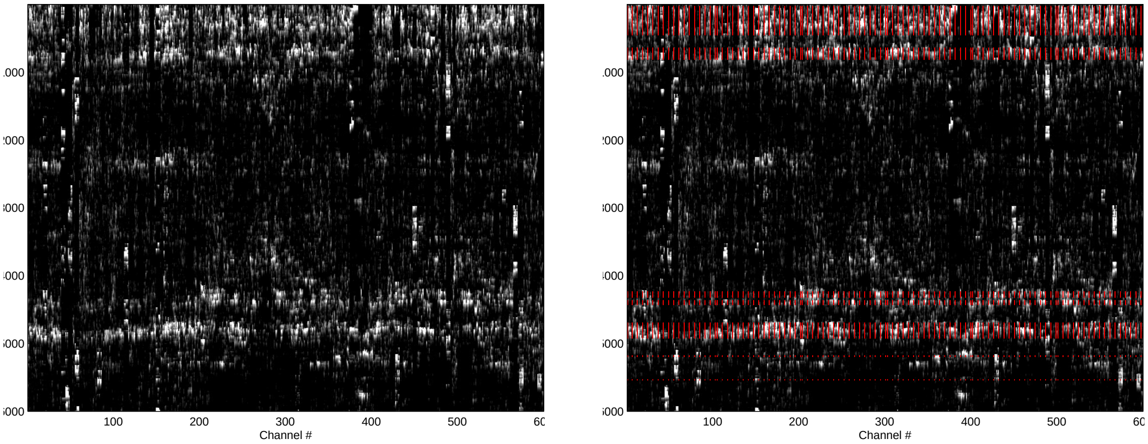
After calculating STA/LTA on the file and storing these values, we apply two threshold-based tests to determine whether events are present. First, we zero all STA/LTA values below the STA/LTA threshold ($\tau_{sta/lta}$) in order to improve the STA/LTA signal-to-noise ratio (i.e., the contrast of events from background noise). As with ℓ_{sta} and ℓ_{lta}, $\tau_{sta/lta}$ must be chosen by trial and error on a short data sample (5-10 minutes). Specifically, a threshold value that is less than the average STA/LTA value of all visually identifiable transportation-related events but greater than the average STA/LTA value of background intervals between events is preferred.

Second, we calculate the square of sums for each time sample on a running time window across all channels. For each file, this yields a single vector containing the short-term energy of the STA/LTA ratio averaged over the all channels at any given time, which we denote $\epsilon_{sta/lta}$. Values of $\epsilon_{sta/lta}$ that exceed an energy threshold τ_ϵ are marked as events, producing a binary event catalog of record length with 1 at time samples where a transportation-related event has been detected and 0 otherwise. The energy threshold τ_ϵ must also be set by experimentation on a small dataset. Because this is the most sensitive parameter for distinguishing true events from both background and optical noise, τ_ϵ performs better when designed on as large a sample dataset as possible. In particular, τ_ϵ may need to change with the balance of traffic and background noise intensities during night time and rush hour in order to perform most effectively. For the Fairbanks dataset, we found $\tau_{sta/lta} = 3$ and $\tau_\epsilon = 500$ to be effective threshold values.

On the Richmond dataset, where the geometry of the array relative to traffic results in no discernible group moveout of vehicle-induced surface waves, the short-term energy sum can be applied with a horizontal window that examines all channels at the same time sample, because the phase moveout of events at apparent surface wave velocities is effectively instantaneous relative to the sample rate $f_s = 1 - 2$ kHz (see Figure 3).

The Fairbanks dataset, however, showcases one of the potential challenges of this detection method because the geometry of the array relative to traffic results in a group moveout at low velocity (approx. 25 m/s). This means that a horizontal window will only ever include a portion of the event, and the sensitivity of $\epsilon_{sta/lta}$ will be reduced because channels of background noise will be averaged with any real event. Additionally, a down-weighting filter designed on the resulting event catalog will remove a significant quantity of usable data. To address this challenge, we correct the STA/LTA matrix for linear moveout (LMO) at the approximate traffic speed before calculating $\epsilon_{sta/lta}$ so that events are parallel to the moving time window and have a stronger signature. This is similar to slant stacking.

Because traffic can be traveling in two directions on any given road, $\epsilon_{sta/lta}$ must be calculated twice with LMO in either direction. We distinguish the favorable direction (which catalog to use, i.e., the direction in which most cars are traveling in any file) using the properties of the energy distribution with time. For a file containing a single car (a unimodal energy distribution with time), the kurtosis of $\epsilon_{sta/lta}$ will be less for the favorable direction because the window will sum across a flat event as compared to an event with significant move out. For files containing multiple cars (a multimodal energy distribution with time), the skewness of $\epsilon_{sta/lta}$ will be greater for the favorable direction for a similar reason. When an LMO correction has been applied during event detection, a direction variable must be passed with the event catalog in order to design a down-weighting filter with the correct



(a)

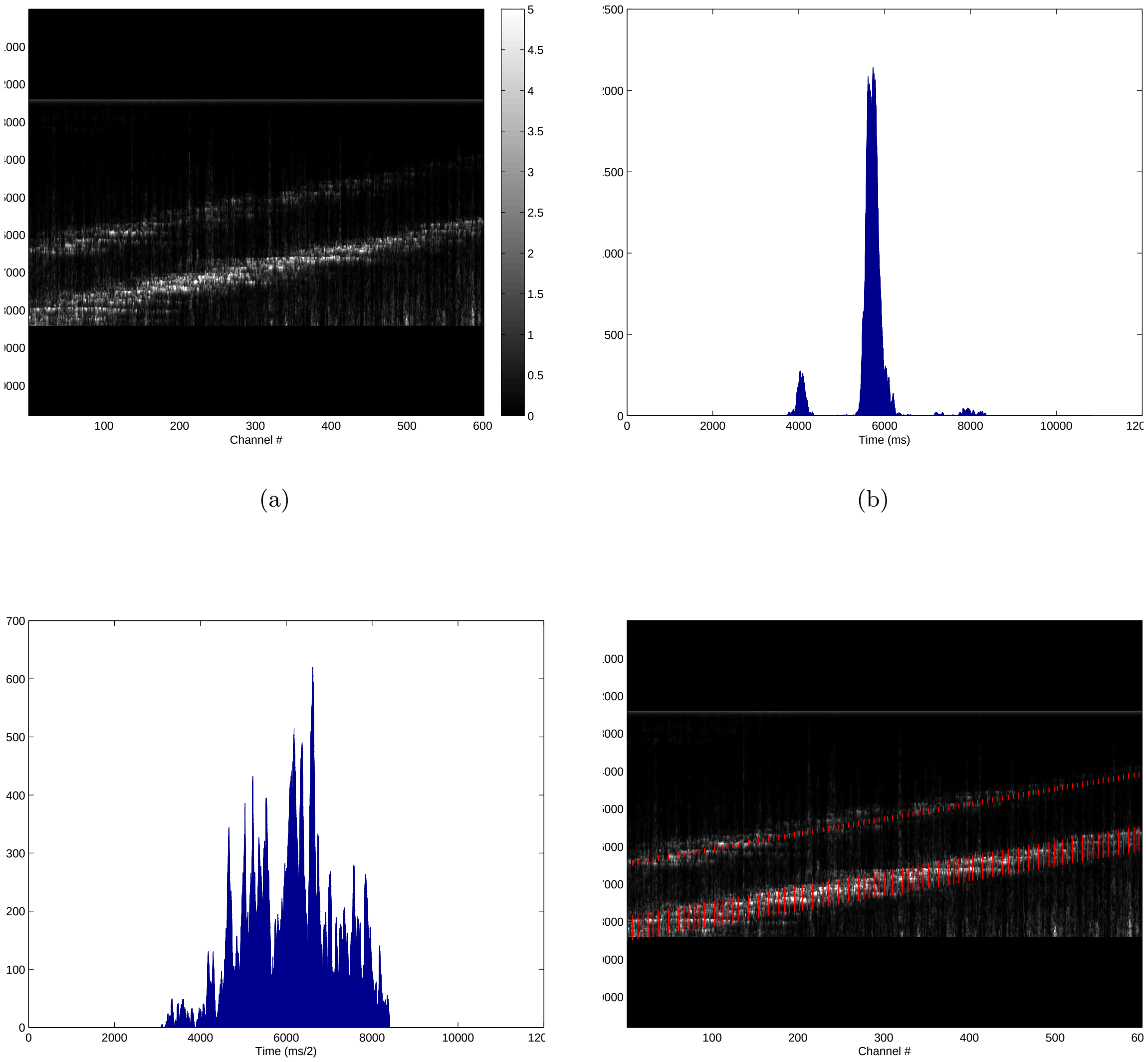
(b)

Figure 3: (a) The STA/LTA image of an example 60 s file from the Richmond dataset. (b) Shading shows the region included in the final catalog, superimposed over (a). [**CR**] ethan1/. richmond-stalta,richmond-energy

shape. If a file records cars traveling in multiple directions at the same time (e.g., Figure 5), this method selects the direction with the strongest events, and either will catalog the entire interval where events in the opposite direction occur or will not successfully identify these events, depending on their relative strength.

Down-weighting

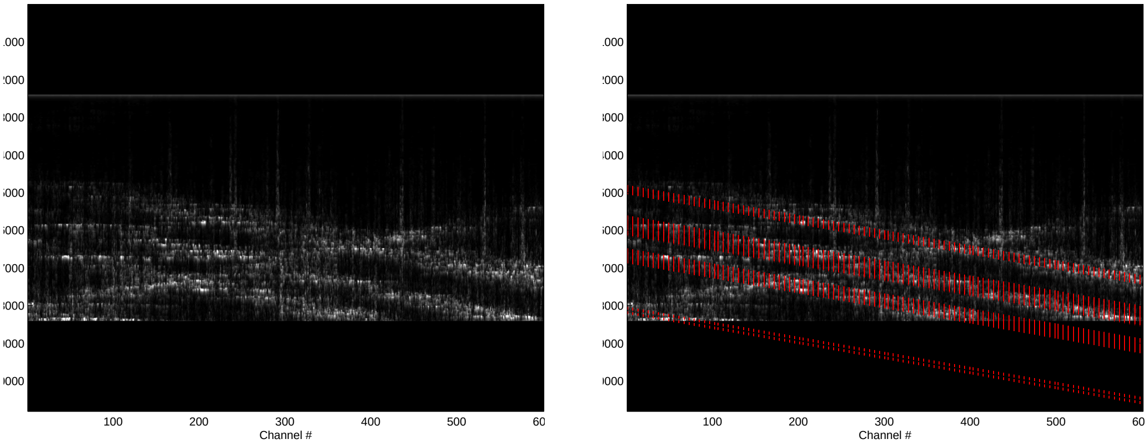
When calculating cross-correlations on the data, we use a simple down-weighting scheme to remove events using the catalogs which were produced by running the tools outlined above in real time as the data were recorded. First, a data file is imported along with its event catalog and basic processing is done, including de-spiking (Martin et al., 2015), temporal normalization, and spectral whitening (Bensen et al., 2007). Second, a filter is applied to the data which is zero at every time sample corresponding to a 1 in the event catalog and one at every time sample corresponding to a 0 in the event catalog, except at the edge of events, where it follows a Gaussian taper. This taper prevents sharp transitions in the down-weighted file and should ensure that imperfect detection parameters which only identify portions of events are still effective at reducing their contribution to the final correlations. When a LMO correction has been applied during event detection, the direction variable associated with the event catalog is used to extrapolate the filter from one channel to the next. Figure 6 shows an example of a down-weighting filter applied to a file from the Fairbanks dataset.



(a) (b)

(c) (d)

Figure 4: (a) The STA/LTA image of an example 60 s file from the Fairbanks dataset. (b) The $\epsilon_{sta/lta}$ calculated with LMO correction in the correct direction. (c) The $\epsilon_{sta/lta}$ calculated with LMO correction in the incorrect direction. (d) Shading shows the region included in the final catalog superimposed over (a). [**CR**] ethan1/. fairbanks-stalta,fairbanks-energy,fairbanks-wrong,fairbanks-catalog



(a)

(b)

Figure 5: (a) The STA/LTA image of an example 60 s file from the Fairbanks dataset where multiple events cross. (b) Shading shows the region included in the final catalog superimposed over (a). [**CR**] ethan1/. fairbanks-badstalta,fairbanks-badcatalog

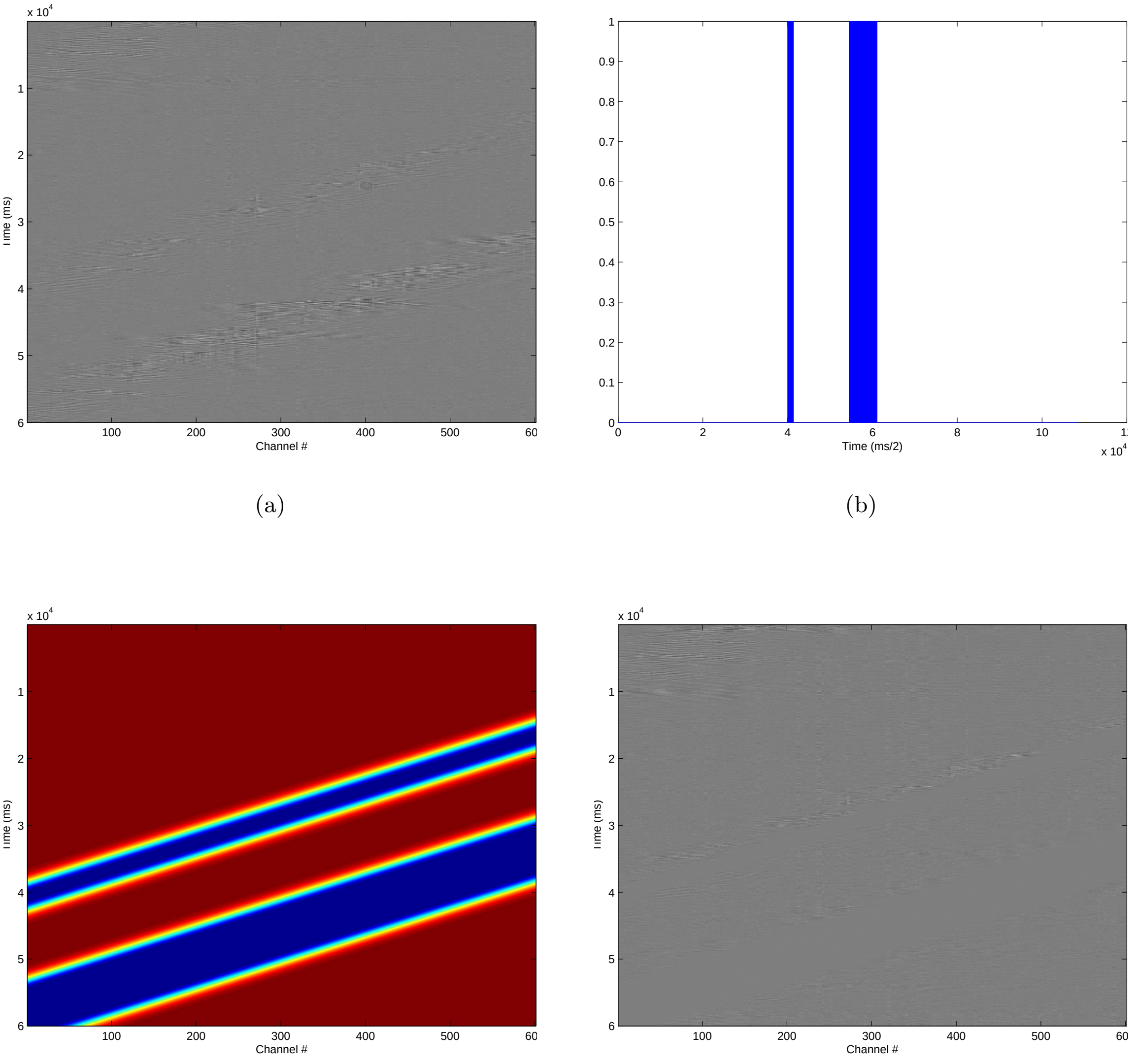
RESULTS ON THE TEST DATA

We calculated cross-correlations on 40 minutes of ambient noise from the Fairbanks dataset following the methods of Martin et al. (2015) and Bensen et al. (2007), with the goal of showing the effectiveness of our proposed processing workflow. Lindsey (2016) showed that manual identification and removal of transportation related noise improved extracted Greens functions significantly. Using a smaller test dataset and less refined pre-processing tools, we were able to show a decrease in the strength of artifacts in cross-correlation images and a small increase in the degree of convergence of the Rayleigh wave fundamental mode associated with the virtual source channel (see Figure 7). More refined workflows, such as in Lindsey (2016), have the potential to improve the image further.

CONCLUSIONS

Overall, the use of an automated processing workflow for identifying and removing coherent anthropogenic noise from ambient noise data recorded by a dense DAS array improved the quality of extracted Greens functions. Our proposed method calculates the STA/LTA ratio on each file, uses a running window check to pick out events, and down-weightes these events in the final correlations. The method is adaptable to different recording geometries, and we present one example of a variation of the method that corrects for apparent group moveout at road speeds.

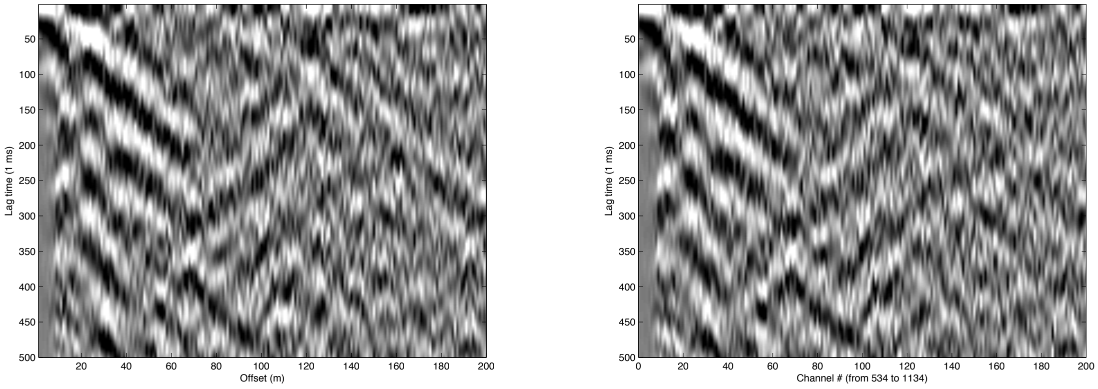
This processing workflow is most effective when applied to recordings with low to moderate levels of traffic-related noise, because it is unable to recover significant high-quality ambient noise data from recordings capturing more than 3-4 vehicles or where vehicles trav-



(a) (b)

(c) (d)

Figure 6: Effect of downweighting on an example 60 s file from Fairbanks. (a) The raw data. (b) The binary catalog determined as above. (c) The downweighting filter designed using the event catalog derived from (b) and the direction of the LMO correction. (d) The data after downweighting. [**CR**] ethan1/. fairbanks-raw,fairbanks-cat,fairbanks-filter,fairbanks-dw



(a) (b)

Figure 7: Cross-correlations on 40 mins of ambient noise data recorded by the trenched DAS array in Fairbanks, AK. The left image (a) was created using the complete recordings, and the right image (b) was created using recordings where transportation-related events had been down-weighted. [**CR**] ethan1/. before,after

eling in two directions frequently cross the array at the same time. The method is able to recover the most usable data from recordings capturing 1-3 vehicles traveling in the same direction or where vehicles traveling in two directions do not cross the array at the same time. This extends the amount of daytime recordings that we are able to use in interferometry without generating strong artifacts. By increasing the fraction of recorded data that can be used in interferometry, we are increasing the potential of this method for time-lapse monitoring of the near-surface because reliable Greens functions can be extracted more frequently, providing a higher temporal resolution in monitoring studies.

Future development of this method has the potential to include a more sophisticated filter design process which is able to down-weight cars that cross paths along the array, as well as the consideration of metrics which determine whether an entire file needs to be thrown out without following the entire processing workflow. This method will be applied in a survey currently taking place on Stanford campus in order to test the potential of ambient noise recorded with DAS to monitor traffic and to develop an event catalog that can be used in future processing work.

ACKNOWLEDGMENTS

We would like to thank Biondo Biondi for many helpful discussions about this project. We would also like to thank the SERDP Permafrost Thaw Experiment team for providing the datasets used in this report. The data collection was supported by the US Department of Defense under SERDP grant RC-2437 "Developing Smart Infrastructure for a Changing Arctic Environment Using Distributed Fiber-Optic Sensing Methods." The team, led by PI Jonathan Ajo-Franklin (Lawrence Berkeley Lab) and co-PI Anna Wagner (US Army

Corps of Engineers CRREL), includes Kevin Bjella (CRREL), Tom Daley (LBL), Shan Dou (LBL), Barry Freifeld (LBL), Nate Lindsey (LBL/UC Berkeley), Michelle Robertson (LBL), and Craig Ulrich (LBL). Eileen Martin's work is supported in part by the Department of Energy Computational Science Graduate Fellowship, provided under grant number DE-FG02-97ER25308. Ethan Williams's work is supported by a grant from the Stanford Earth Summer Undergraduate Research (SESUR) Program.

REFERENCES

Ajo-Franklin, J., N. Lindsey, S. Dou, T. M. Daley, B. M. Freifeld, E. R. Martin, M. Robertson, C. Ulrich, and A. Wagner, 2015, A field test of distributed acoustic sensing for ambient noise recording. SEG Extended Abstracts.

Bensen, G. D., M. H. Ritzwoller, M. P. Barmin, A. L. Levshin, F. Lin, M. P. Moschetti, N. M. Shapiro, and Y. Y, 2007, Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements: Geophysical Journal International, **169**, 1239–1260.

Daley, T. M., B. M. Freifeld, J. Ajo-Franklin, and S. Dou, 2013, Field testing of fiber-optic distributed acoustic sensing (das) for subsurface seismic monitoring: The Leading Edge, **June 2013**, 699–706.

Lindsey, N., 2016, Seismic interferomerty with distributed acoustic sensing for near-surface monitoring of critical infrastructure. Seismological Society of America.

Martin, E. R., J. Ajo-Franklin, S. Dou, N. Lindsey, T. M. Daley, B. M. Freifeld, M. Robertson, A. Wagner, and C. Ulrich, 2015, Interferometry of ambient noise from a trenched distributed acoustic sensing array. SEG Extended Abstracts.

Martin, E. R., N. J. Lindsey, S. Dou, J. B. Ajo-Franklin, A. Wagner, K. Bjella, T. M. Daley, B. Freifeld, M. Robertson, and C. Ulrich, 2016, Interferometry of a roadside das array in fairbanks, ak. SEG Extended Abstracts (to appear).

Nakata, N., R. Snieder, T. Tsuji, K. Larner, and T. Matsuoka, 2011, Shear wave imaging from traffic noise using seismic interferometry by cross-coherence: Geophysics, **76**, SA97–SA106.

Wapenaar, K., D. Draganov, R. Snieder, X. Campman, and A. Verdel, 2010, Tutorial on seismic interferometry: Part 1 – basic principles and applications: Geophysics, **75**, 75A195–75A209.

Withers, M., R. Aster, C. Young, J. Beiriger, M. Harris, S. Moore, and J. Trujillo, 1998, A comparison of select trigger algorithms for automated global seismic phase and event detection: Bulletin of the Seismological Society of America, **88**, 95–106.

Yoon, C. E., O. O'Reilly, K. J. Bergen, and G. C. Beroza, 2015, Earthquake detection through computationally efficient similarity search: Science Advances, **1**, e1501057–e1501057.

Stanford Exploration Project, SEP165, October 8, 2016

Exploratory data analysis of the SOLA land dataset

Fantine Huot

ABSTRACT

Seismic surveys provide us with an abundance of data characteristics. Is there an informative way to visualize a survey's metadata? Using exploratory data analysis (EDA) tools, we investigate a land survey's metadata to detect trends among the observations. We derive indicators of the quality of a seismic trace.

INTRODUCTION

With the advent of multi-component sensors, we are faced with an abundance of data characteristics which may contain precious information about our survey. For example, within a seismic survey, noise levels vary by receiver location, source location, offset, frequency, time of the day, etc. While bad traces are easy to pick out, the study of data characteristics can yield criteria for assessing the quality of a seismic trace.

High noise levels degrade seismic images. Thus, significant efforts are deployed in seismic signal processing to attenuate and remove noise from the signal (Yilmaz, 2001), while stacking techniques are commonly used to improve signal to noise ratio (Claerbout and Black, 2008). Once bad traces are identified, they can be selectively left out of the imaging process using weighing and penalty functions as introduced by Claerbout (2014).

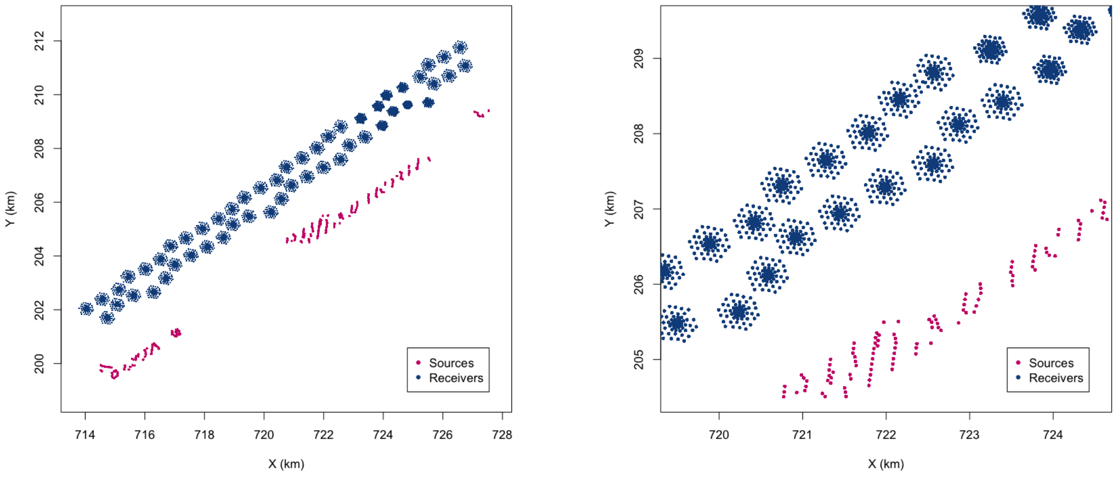
Therefore, it would be extremely useful to derive indicators of noise level or recording quality from a survey's metadata. However, what would be an informative way to visualize these metadata? How can we efficiently discover trends among the variables or among the observations? Which are the important variables?

Exploratory data analysis (EDA) refers to a diverse set of techniques for answering questions such as these. Herein we use R, a free software environment for statistical computing, to perform our analysis. We conduct our study on the metadata of the SOLA dataset.

THE SOLA LAND DATASET

Acquisition geometry

The SOLA survey is a three-component 2D array land acquisition conducted during summer 2015. It has 3,600 receivers arranged into 54 hexagonal arrays and 4,000 shots. The arrays are 200 m to 600 m wide, with 63 to 110 receivers per array. In order to record a wide range of frequencies while still providing comprehensive coverage, the arrays are irregularly sampled: the receiver spacing varies from 5 to 10 m on the inside of the array to 50 m on the outside. The survey has about 7 million traces, but for this study we were provided with the data characteristics of a subset of 350,000 traces. This subset's acquisition geometry is illustrated in Figure 1.



(a) (b)

Figure 1: (a) The SOLA acquisition geometry. (b) Zoomed in portion of the survey. In order to record a wide range of frequencies while still providing comprehensive coverage, the receivers are arranged in irregularly sampled hexagonal arrays. [**ER**] fantine2/. geom1,geom2

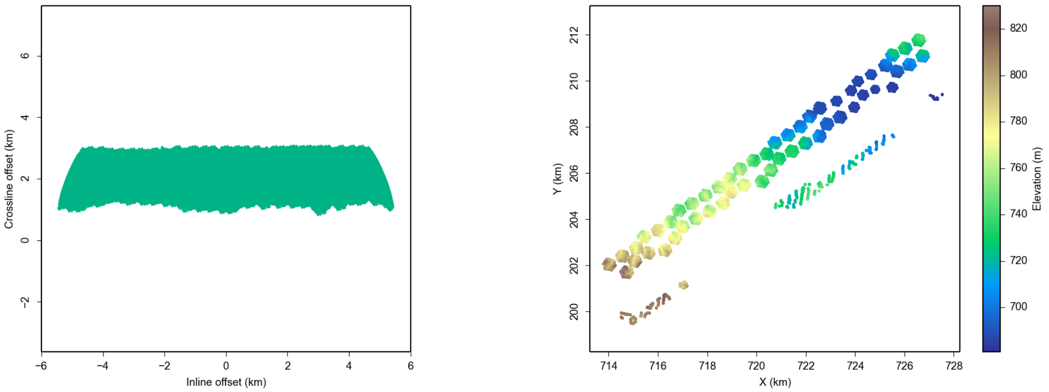
The metadata we will be considering for this study contains two types of entries: descriptive headers and data characteristics.

Descriptive headers

The descriptive headers cover all acquisition parameters for each seismic trace, such as source and receiver location, shot times, offset and azimuth (Figure 2). The complete list of headers is provided for reference in the appendix. Certain headers were left out from this study for they were redundant or did not vary over the selected subset, leaving us with 25 variables.

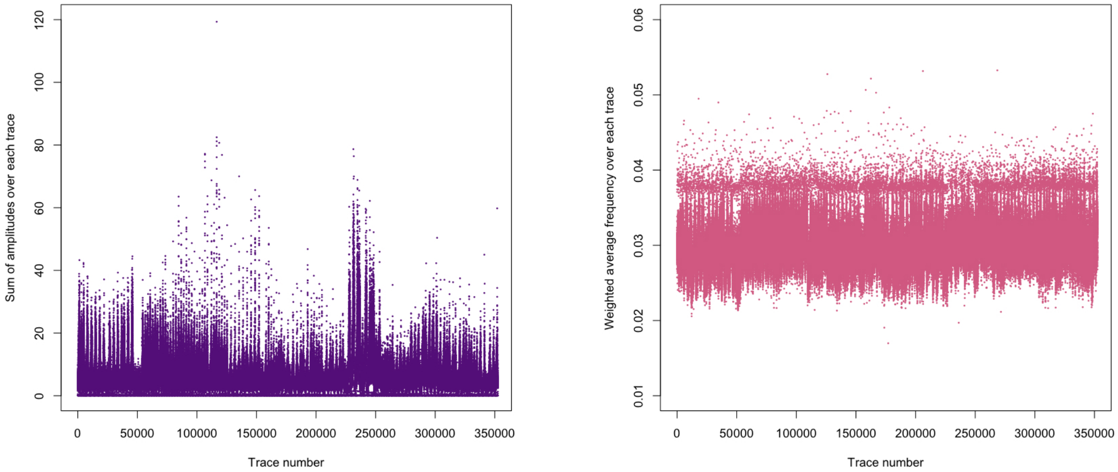
Data characteristics

The data quality varied greatly over the survey, and the acquisition recorded various levels of surface noise. To capture this variability, a certain number of data characteristics were computed for each trace. These data characteristics include variables such as first break pick, average amplitude, and average frequency or spikiness of the signal (Figure 3). When possible, they were computed both over the full trace and over various time windows, constituting a total of 52 variables. The complete list of computed data characteristics is provided in the appendix.



(a) (b)

Figure 2: Examples of parameters contained in the survey's descriptive headers. (a) Inline and crossline offset. (b) Source and receiver elevation. [**ER**] fantine2/. header1,header2



(a) (b)

Figure 3: Examples of the computed data characteristics. (a) Sum of amplitudes over each trace. (b) Weighted average frequency of each trace. [**ER**] fantine2/. data1,data2

Using both the header information and the computed data characteristics, we herein seek to identify trends in the noise variability.

DATA VISUALIZATION

Combining the descriptive headers and data characteristics, we have a total of $n = 350,000$ observations over $p = 77$ different variables. We could visualize these data by plotting two-dimensional scatter plots, each of which contains the n observations' measurements on two of the p variables. However, there are $\binom{p}{2} = p(p-1)/2 = 2,926$ such scatter plots, which makes it prohibitive to look at all of them. Moreover, most of them would not be very informative since they each contain only a small fraction of the total information present in the dataset.

Principal component analysis (PCA)

When faced with a large set of correlated variables, principal components allow us to summarize the dataset with a smaller number of representative variables that collectively explain most of the variability in the original set. The idea is that each of the n observations lives in a p-dimensional space, but not all of these dimensions are equally interesting. Principal component analysis (PCA) seeks a small number of dimensions that are as interesting as possible, where the concept of interesting is measured by the amount that the observations vary along each dimension. The dimensions found by PCA are called principal components. Principal components are a sequence of linear combinations of the p variables, mutually uncorrelated and ordered in variance. They are the directions along which the original data is highly variable.

In the following, we provide a brief overview on how to compute the principal components of a dataset, based on formulations by Hastie et al. (2005). The first principal component Z_1 of a set of variables $X_1, X_2, ..., X_p$ is the normalized linear combination,

$$Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + ... + \phi_{p1}X_p,$$

that has the largest variance. The elements $\phi_{11}, ..., \phi_{p1}$ are called the loadings of the first principal component. Together, they make up the principal component loading vector, $\phi_1 = (\phi_{11}\,\phi_{21}\,...\,\phi_{p1})^T$. As an arbitrarily large value of these loadings would result in an arbitrarily large variance, the loadings are normalized such that $\sum_{j=1}^{p} \phi_{j1}^2 = 1$.

Let's consider a certain $n \times p$ dataset $\mathbf{X}$. As we are only interested in variance, we assume that each of the variables in $\mathbf{X}$ has been centered to have mean zero. The first principal component of $\mathbf{X}$ is computed by finding the linear combination of the sample variable values of the form $z_{i1} = \phi_{11}x_{i1} + \phi_{21}x_{i2} + ... + \phi_{p1}x_{ip}$, that has largest sample variance, subject to the constraint that $\sum_{j=1}^{p} \phi_{j1}^2 = 1$. In other words, the first principal component loading vector solves the following optimization problem:

$$\max_{\phi_{11},...,\phi_{p1}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{p} \phi_{j1} x_{ij} \right)^2 \right\} \text{subject to} \sum_{j=1}^{p} \phi_{j1}^2 = 1. \tag{1}$$

The objective in Equation 1 can be re-expressed as $\frac{1}{n}\sum_{i=1}^{n} z_{i1}^2$. Since $\frac{1}{n}\sum_{i=1}^{n} x_{ij} = 0$, the average of the $z_{11}, ..., z_{n1}$ will be zero as well. Hence the objective that we are maximizing in Equation 1 is just the sample variance of the n values of z_{i1}. Equation 1 can then be solved via an eigen decomposition (Golub and Van Loan, 1983).

Once the first principal component Z_1 has been determined, the second principal component is the linear combination of $X_1, X_2, ..., X_p$ that has maximal variance out of all the linear combinations that are uncorrelated with Z_1. Constraining Z_2 to be uncorrelated with Z_1 is equivalent to constraining the loading vector ϕ_2 to be orthogonal to ϕ_1. As a consequence, to find the second principal component, we solve a problem similar to the one expressed in Equation 1, with ϕ_2 replacing ϕ_1, and with the additional constraint that ϕ_2 is orthogonal to ϕ_1.

Data projection

By projecting the data along the first few principal component directions, we can build two-dimensional representations that capture most the dataset's variability. PCA was performed on the SOLA metadata after standardizing each variable to have zero mean and standard deviation one. Figure 4 represents the dataset projected along its three first principal components.

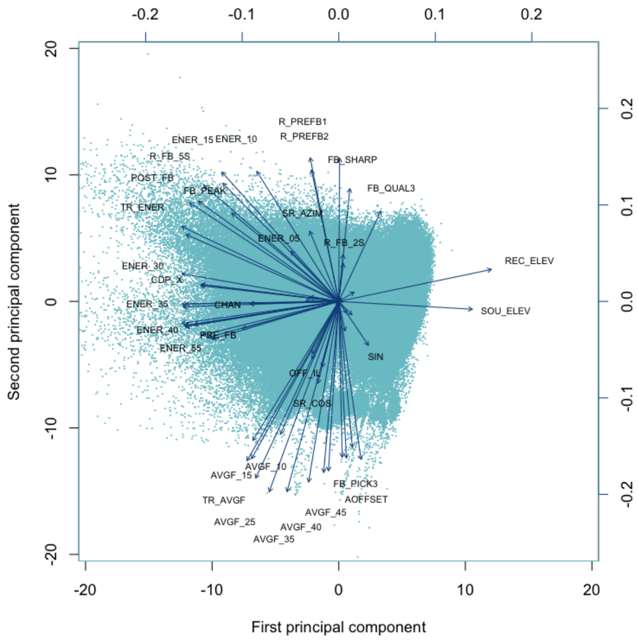
This representation allows us to visualize the nature of the first principal components. The first principal component puts weight mostly on source and receiver elevation, common depth point location, and average energy over various time windows. The second principal component accounts for absolute offset, first break pick and average frequency over various time windows. The variability over the seismic traces seems to be mostly explained in terms of source and receiver elevation and absolute offset as far as the descriptive headers are concerned, and average energy and frequency content for the data characteristics. Among the various time windows on which the summed energy and average frenquencies were computed, 2.8 to 3.2 s seems to carry most weight. The variables that accounted for the spikiness of the data have little to no impact on the first principal components, and only start carrying weight from the 11th component onwards.

Proportion of variance explained

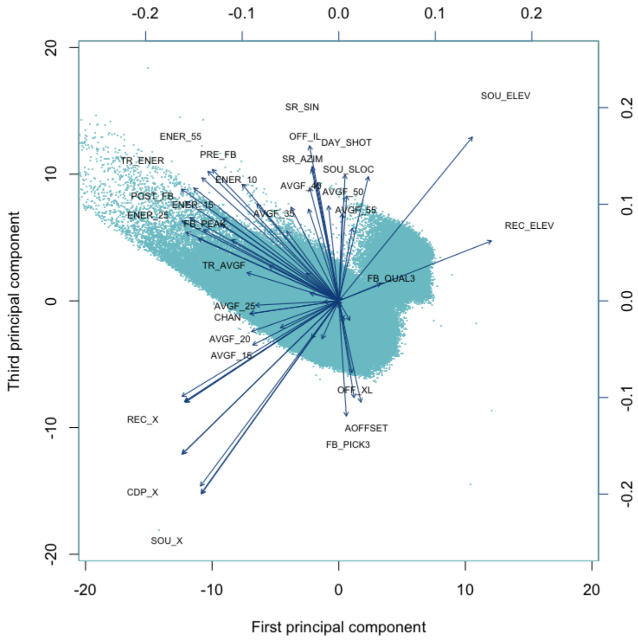
Although there are a possible $p = 77$ principal components, approximately 23 account for 90% of the total variation, while the first three ones account for 46%. Together, the first seven principal components explain around 65% of the variance in the data. This may not seem a large amount of variance. However, from the plots in Figure 5, we see that while each of the first seven principal components explain a substantial amount of variance, there is a marked decrease in the variance explained by further principal components. This suggests that there may be little benefit to examining more than seven principal components.

IDENTIFYING TRENDS IN THE DATA

Variables that are located close to each other in the PCA projection indicate potential correlations, enabling us to identify trends in the data. Each panel of Figures 6, 7, 8 and 9 is a scatterplot for a pair of variables whose identities are given by the corresponding row and column labels.

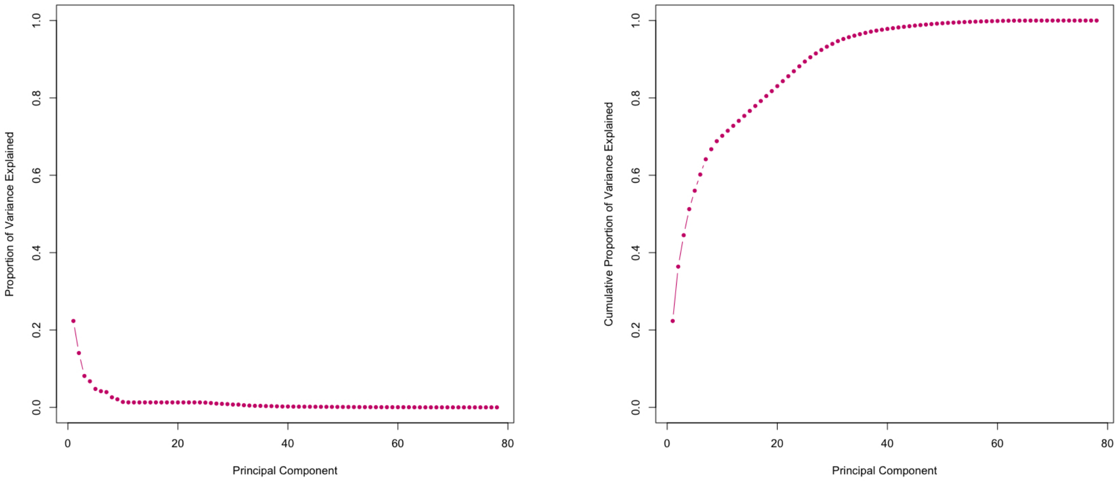


(a)



(b)

Figure 4: Projection of the SOLA metadata on its first principal component directions. The dots represent the projected metadata, while the arrows indicate the loadings associated to each variable. For readability reasons, only the loadings greater than 0.05 are labeled. The description of each label is provided in the appendix. (a) Projection on the first and second principal components. (b) Projection on the first and third principal components. [**CR**] fantine2/. pca1,pca2



(a) (b)

Figure 5: (a) Proportion of variance explained by each principal component. (b) Cumulative proportion of variance explained. [**CR**] fantine2/. pve1,pve2

Amplitudes seem to decrease with receiver elevation (Figure 6). The average frequency variation narrows down with receiver elevation (Figure 7). As to be expected, the time of first break is highly correlated with the absolute offset (Figure 8). However, the decrease of energy with offset only shows on the variables computed on early time windows. Unsurprisingly, the quality of first break pick is closely linked to the amplitudes before and after first break, and the first break's maximum amplitude (Figure 9).

IDENTIFYING NOISY TRACES

Using the data characteristics computed on various time windows, we can identify portions of the signal which deviate from the remainder of the signal by unacceptable amounts, as these are likely to correspond to noise bursts.

In order to flag outliers, we use the interquartile range (IQR), which is the difference between the upper and lower quartiles of the data ($IQR = Q_3 - Q_1$), as a measure of statistical dispersion. According to Tukey's range test (Tukey, 1977), outliers are observations that fall outside the range:

$$\big[Q_1 - kIQR, Q_3 + kIQR\big],$$

where k is a positive constant. In our study, we use $k = 1.5$, a value commonly used in statistics. On a normal distribution, this value flags less than 1% of the data as outliers.

Figure 10 plots the outliers on the data characteristics both in PCA projection and over the survey's acquisition geometry. It appears that certain receiver locations accounted for noisy measurements. For comparison, Figure 11 represents the same plot, but only marks the outliers associated with spikiness of the signal.

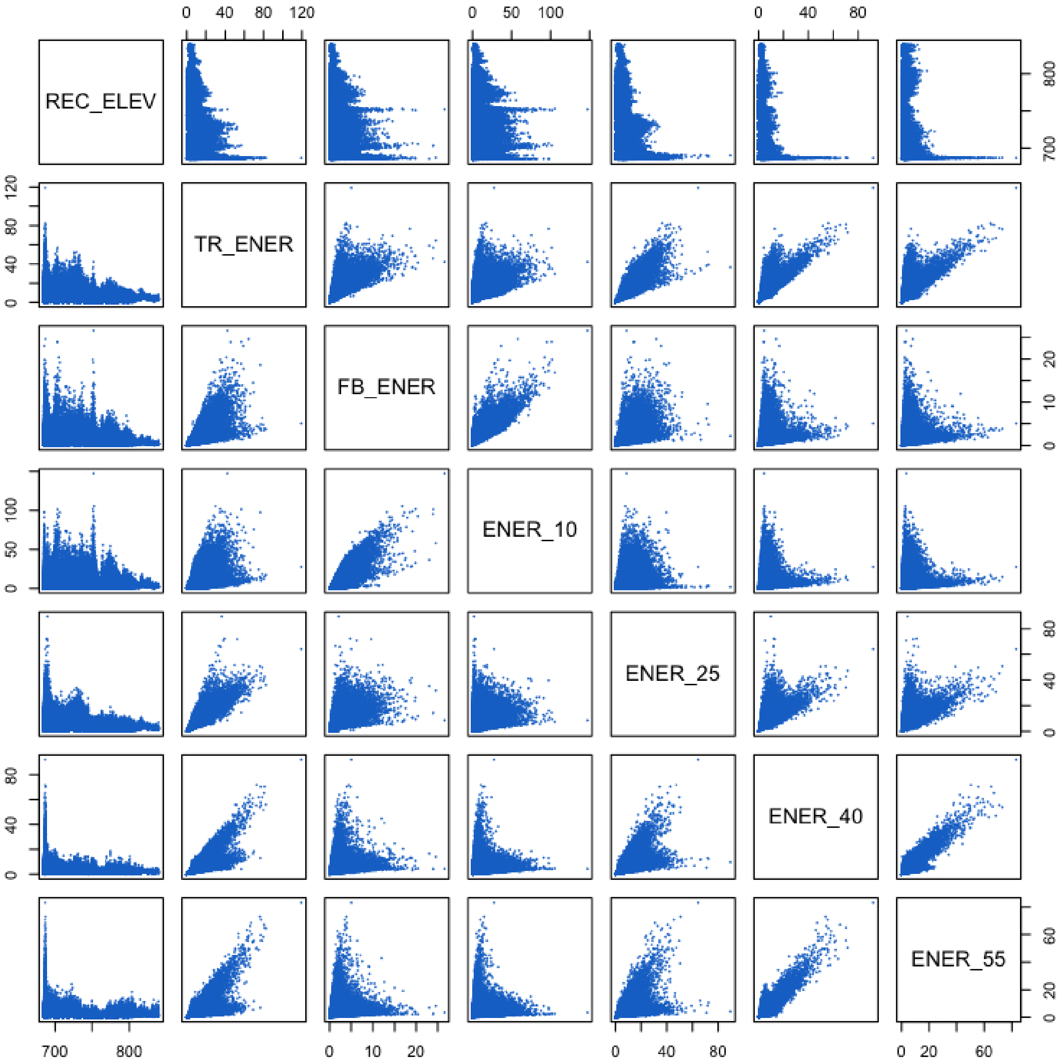


Figure 6: Scatterplots of receiver elevation (REC_ELEV) versus various measures of energy over different time windows. The full description of each variable is provided in the appendix. We notice that amplitudes seem to decrease with receiver elevation. [**CR**] fantine2/. scatter1

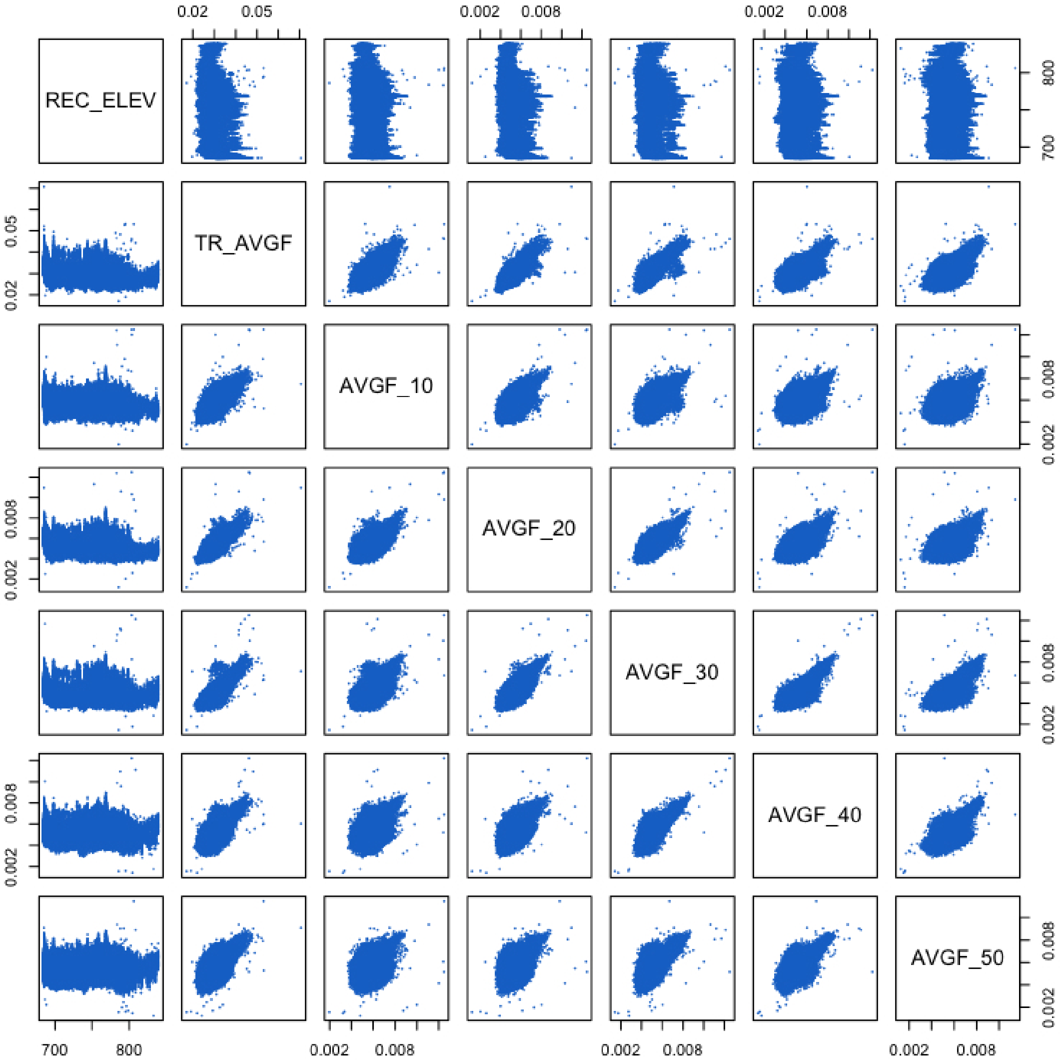


Figure 7: Scatterplots of receiver elevation (REC_ELEV) versus various measures of average frequency over different time windows. The full description of each variable is provided in the appendix. We notice that average frequency variation narrows down with receiver elevation. [**CR**] fantine2/. scatter2

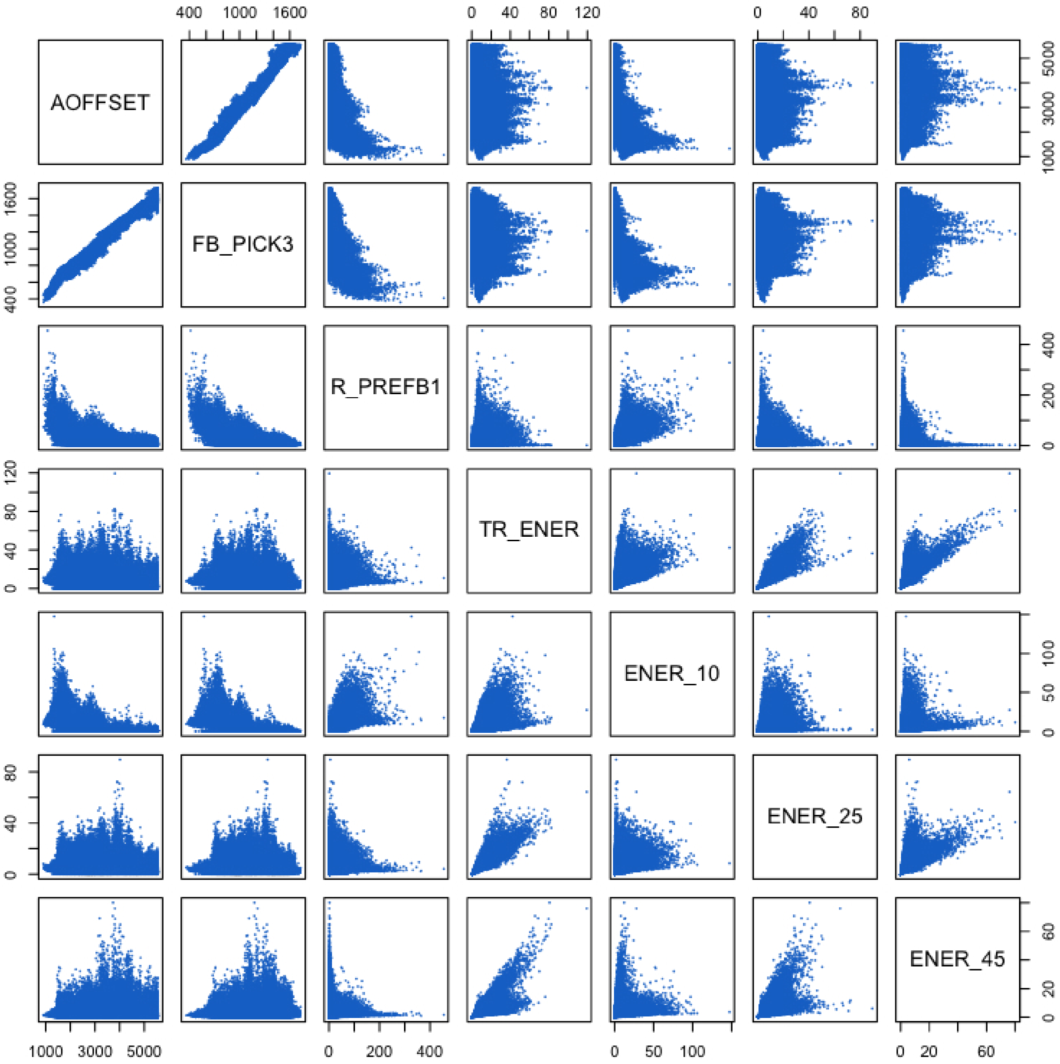


Figure 8: Scatterplots of absolute offset (AOFFSET) versus first break pick (FB_PICK3) and various measures of energy over different time windows. The full description of each variable is provided in the appendix. Unsurprisingly, we notice that first break pick is highly correlated with absolute offset. However, the decrease of energy with offset only shows on the variables computed on early time windows. [**CR**] fantine2/. scatter3

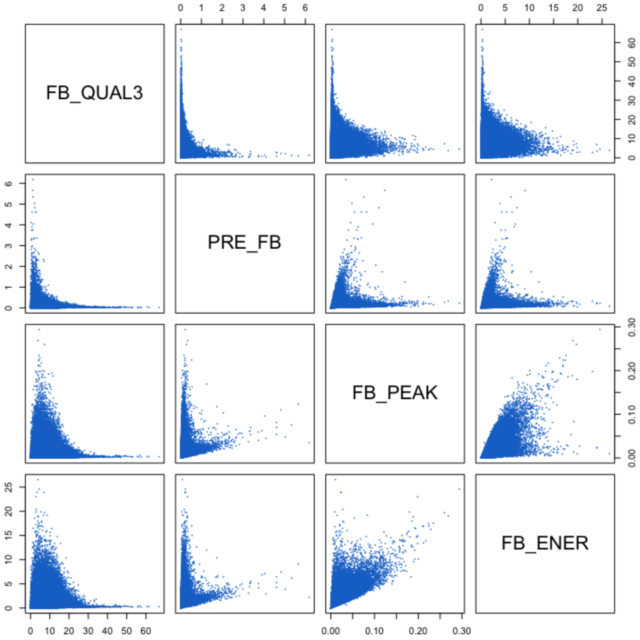
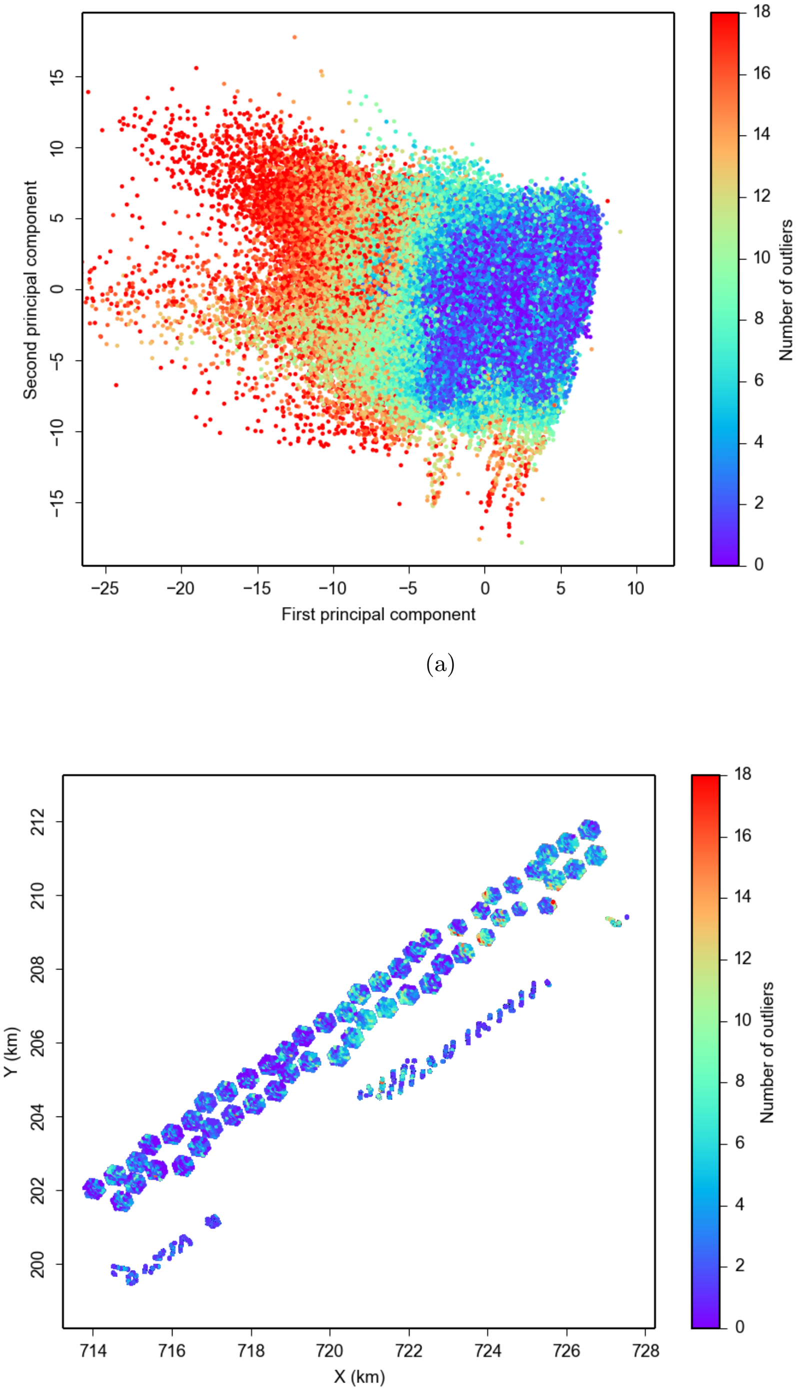


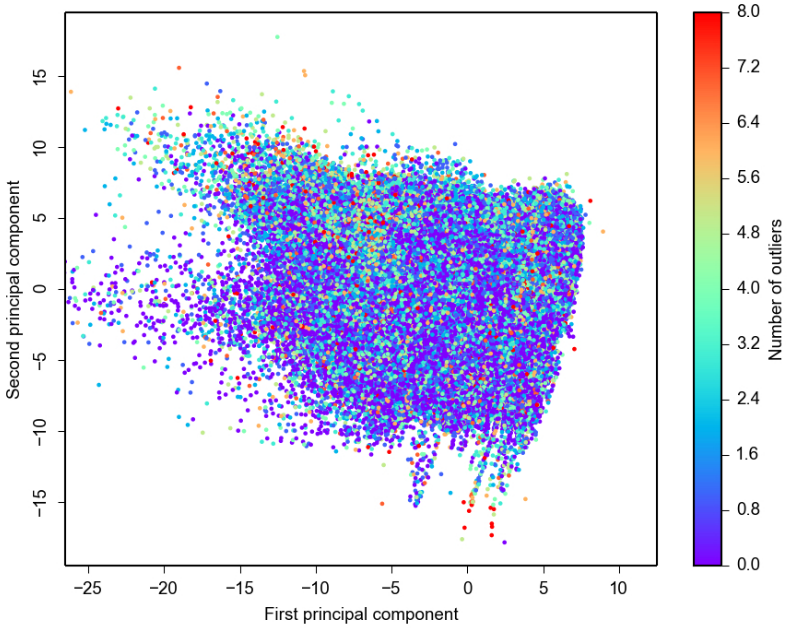
Figure 9: Scatterplots of quality of first break pick (FB_QUAL3) versus amplitudes before (PRE_FB) and after (FB_ENER) first break, and the first break's maximum amplitude (FB_PEAK). Unsurprisingly, all these variables are closely linked. The full description of each variable is provided in the appendix. [**CR**] fantine2/. scatter4



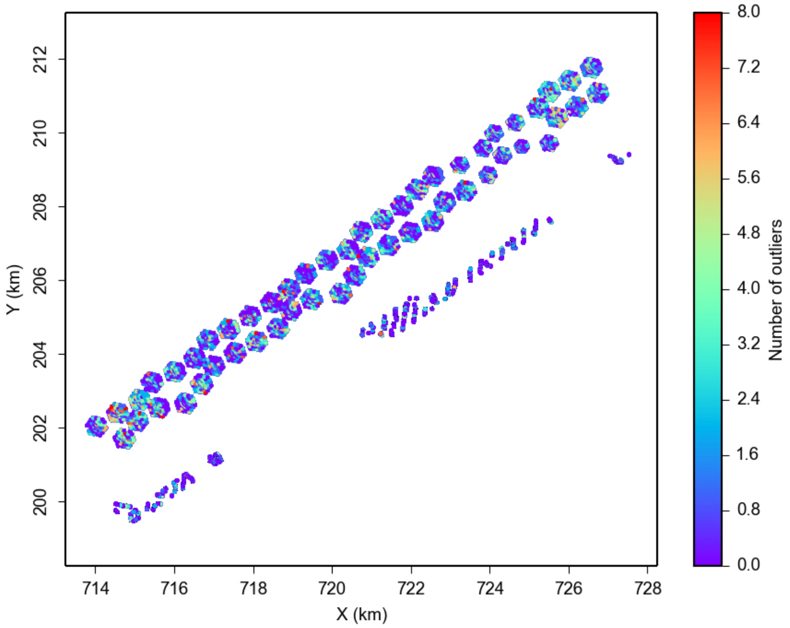
(a)

(b)

Figure 10: Outliers identified by Tukey's range test for $k = 1.5$. (a) In PCA projection. (b) Over the survey's acquisition geometry. Certain receiver locations accounted for noisy measurements. [**CR**] fantine2/. outlier1,outlier2



(a)



(b)

Figure 11: Outliers identified by Tukey's range test for $k = 1.5$ for all the variables related to spikiness of the signal. (a) In PCA projection. (b) Over the survey's acquisition geometry. [**CR**] fantine2/. spikiness1,spikiness2

STATISTICAL ROBUSTNESS

For statistical soundness, each operation presented in this study was performed on ten different subsets of the data, where each subset contained 90% of the original data sampled at random. The different subsets did not show any significant change in the trends presented here.

DISCUSSION

Statistical computing provides useful tools for exploring new datasets and searching for outliers and trends. Free software environments such as R ease the implementation burden for scientists and practitioners. The operations conducted in this study reflect some basic steps commonly used in exploratory data analysis. While the results obtained may not seem surprising to the trained geoscientist, a deeper study may yield more intricate trends. Therefore, it would be of interest to extend this study to the full scope of the original 7 million traces.

Moreover, a statistical approach may help identify and quantify different types of noise by flagging traces that deviate significantly from the remainder of the signal. By extending this study to a larger set of traces, it would be possible to visualize whether noise levels vary in this survey by time of the day, or day of the week. Clustering techniques may help identifying different type of noise.

ACKNOWLEDGMENTS

We would like to thank Christof Stork for providing the data characteristics used in this study.

REFERENCES

Claerbout, J., 2014, Geophysical image estimation by example: Lulu.com.

Claerbout, J. and J. L. Black, 2008, Basic earth imaging: Citeseer.

Golub, G. H. and C. F. Van Loan, 1983, Matrix computations.

Hastie, T., R. Tibshirani, J. Friedman, and J. Franklin, 2005, The elements of statistical learning: Data mining, inference and prediction: 534–552.

Tukey, J. W., 1977, Exploratory data analysis.

Yilmaz, Ö., 2001, Seismic data analysis: Processing, inversion, and interpretation of seismic data: Society of Exploration Geophysicists.

APPENDIX

The complete list of descriptive headers is provided below. The headers marked with a * symbol were left out from this study for they were redundant or did not vary over the selected subset, leaving us with 25 header variables.

SOU_X	Source X coordinate
SOU_Y	Source Y coordinate
SOU_ELEV	Source elevation
DEPTH*	Source depth
UPHOLE*	Source uphole time
SIN	Source internal index number
SOU_SLOC	External source location number
FFID*	Field file index number
SOURCE*	Live source number
S_LINE	Swath or sail line number
REC_X	Receiver X coordinate
REC_Y	Receiver Y coordinate
REC_ELEV	Receiver elevation
GEO_COMP*	Geophone component (x,y,z)
CHAN	Recording channel number
REC_SLOC	Receiver index number
SRF_SLOC	External receiver location number
R_LINE	Receiver line number
CDP_X	X coordinate of common depth point
CDP_Y	Y coordinate of common depth point
OFFSET	Signed source-receiver offset
AOFFSET	Absolute value of offset
OFF_IL	Inline offset
OFF_XL	Crossline offset
SR_AZIM	Source-receiver azimuth
SR_COS	Cosine of source-receiver azimuth
SR_SIN	Sine of source-receiver azimuth
ARRAY_ID	Array index number
YER_SHOT*	Year
DAY_SHOT	Day
TIM_SHOT	Time of the day
TRC_TYPE*	Trace type

The complete list of the 52 computed data characteristics is provided below.

FB_PICK3	First break pick
FB_QUAL3	Quality of fine tuned first break
FB_NDIFF	FB_PICK3 - FB_NAVG
TR_ENER	Sum of amplitudes of trace
TR_AVGF	Weighted average frequency of trace
PRE_FB	Pre-first break energy, normalized
FB_PEAK	Max amplitude of first break
FB_ENER	Energy from first break to 0.5 s afterwards
FB_SHARP	First break sharpness
POST_FB	Energy from 0.5 to 2.0 s after first break
R_PREFB1	FB_ENER / PRE_FB
R_PREFB2	POST_FB / PRE_FB
SPIKT_AL	Spikiness of time data
SPIKF_AL	Spikiness of frequencies
ENER_05	Average amplitude from 0.2 to 0.8 s
ENER_10	Average amplitude from 0.8 to 1.2 s
ENER_15	Average amplitude from 1.2 to 1.8 s
ENER_20	Average amplitude from 1.8 to 2.2 s
ENER_25	Average amplitude from 2.2 to 2.8 s
ENER_30	Average amplitude from 2.8 to 3.2 s
ENER_35	Average amplitude from 3.2 to 3.8 s
ENER_40	Average amplitude from 3.8 to 4.2 s
ENER_45	Average amplitude from 4.2 to 4.8 s
ENER_50	Average amplitude from 4.8 to 5.2 s
ENER_55	Average amplitude from 5.2 to 5.8 s
AVGF_05	Average frequency from 0.2 to 0.8 s
AVGF_10	Average frequency from 0.8 to 1.2 s
AVGF_15	Average frequency from 1.2 to 1.8 s
AVGF_20	Average frequency from 1.8 to 2.2 s
AVGF_25	Average frequency from 2.2 to 2.8 s
AVGF_30	Average frequency from 2.8 to 3.2 s
AVGF_35	Average frequency from 3.2 to 3.8 s
AVGF_40	Average frequency from 3.8 to 4.2 s
AVGF_45	Average frequency from 4.2 to 4.8 s
AVGF_50	Average frequency from 4.8 to 5.2 s
AVGF_55	Average frequency from 5.2 to 5.8 s

SPIKF05	Spikiness of frequencies from 0.2 to 0.8 s
SPIKF10	Spikiness of frequencies from 0.8 to 1.2 s
SPIKF15	Spikiness of frequencies from 1.2 to 1.8 s
SPIKF20	Spikiness of frequencies from 1.8 to 2.2 s
SPIKF25	Spikiness of frequencies from 2.2 to 2.8 s
SPIKF30	Spikiness of frequencies from 2.8 to 3.2 s
SPIKF35	Spikiness of frequencies from 3.2 to 3.8 s
SPIKF40	Spikiness of frequencies from 3.8 to 4.2 s
SPIKF45	Spikiness of frequencies from 4.2 to 4.8 s
SPIKF50	Spikiness of frequencies from 4.8 to 5.2 s
SPIKF55	Spikiness of frequencies from 5.2 to 5.8 s
R_FB_1S	Ratio of energy from 0.5-1.0 s after first break to 1.5-2.0 s
R_FB_2S	Ratio of energy from 0.5-1.0 s after first break to 2.5-3.0 s
R_FB_3S	Ratio of energy from 0.5-1.0 s after first break to 3.5-4.0 s
R_FB_4S	Ratio of energy from 0.5-1.0 s after first break to 4.5-5.0 s
R_FB_5S	Ratio of energy from 0.5-1.0 s after first break to 5.5-6.0 s

Overview of a C++ Helmholtz solver library

Joseph Jennings and John Washbourne

ABSTRACT

SEP received a Helmholtz solver during a visit from Chevron researchers. We describe a high-level overview of the workings of the solver and provide simple examples of how this solver may be used for performing full-waveform inversion.

INTRODUCTION

Since the onset of research on full-waveform inversion (FWI) at SEP, the majority of our implementations have been performed in the time domain (Mora, 1987; Shen, 2014; Biondi and Almomin, 2014). Time-domain methods have the advantage that they are relatively straightforward to implement with explicit finite differences and thus require relatively less memory (Wang et al., 2011). In comparison, frequency domain techniques that solve the Helmholtz equation require factorizing a large, asymmetric, sparse matrix (the Helmholtz operator). In spite of this, frequency-domain methods offer the advantage that they solve the wave-equation for a single frequency, easily allowing for frequency continuation methods that have been used to mitigate cycle-skipping (Bunks et al., 1995). Computationally, they also lend themselves to multiple-source modeling as the factorized matrix can be reused for additional source vectors (Marfurt, 1984). In turn, this property becomes very useful for iterative inversion schemes such as Gauss-Newton and Full-Newton algorithms that require many forward propagations for primary and secondary sources at each iteration (Pratt et al., 1998). For these and other reasons, frequency-domain FWI is becoming more common in the geophysical exploration industry (Washbourne et al., 2013).

As part of the Chevron Center of Research Excellence (CoRE) at Stanford, Chevron researchers visited SEP this summer and shared a Helmholtz solver with SEP researchers. The solver is written in C++ in an object-oriented fashion and utilizes the UMFPACK factorization libraries from SuiteSparse (Davis et al., 2014). The solver consists of three operators: a non-linear forward operator for solving the non-linear wave equation, a linearized operator that maps a perturbation in velocity to a perturbation in the data, and lastly the adjoint of the linearized operator.

In this report, we first give a brief introduction to numerically solving the Helmholtz equation. We then describe the different operators of the Helmholtz solver. With each operator, we provide a simple example for how it might be used in an FWI context.

DISCRETIZATION OF THE HELMHOLTZ EQUATION

We begin with the constant-density acoustic wave equation,

$$\frac{1}{c(\mathbf{x})^2}\frac{\partial^2 u(\mathbf{x},t)}{\partial t^2} - \nabla^2 u(\mathbf{x},t) = f(\mathbf{x},t). \tag{1}$$

If we then take the temporal Fourier transform of equation 1, we arrive at the well-known Helmholtz equation,

$$-\frac{\omega^2}{c(\mathbf{x})^2}u(\mathbf{x},\omega) - \nabla^2 u(\mathbf{x},\omega) = f(\mathbf{x},\omega).$$

Discretizing the Laplacian with second-order finite differences, we write the discrete Helmholtz equation as:

$$\frac{\omega^2}{c_{i,j}^2}u(\omega)_{i,j}+\left[\frac{u(\omega)_{i-1,j}-2u(\omega)_{i,j}+u(\omega)_{i+1,j}}{\Delta x^2}\right]+\left[\frac{u(\omega)_{i,j-1}-2u(\omega)_{i,j}+u(\omega)_{i,j+1}}{\Delta z^2}\right]=-f_{i,j}(\omega).$$

Grouping terms by spatial index, we can write the system of equations as a matrix vector product:

$$\mathbf{A}\mathbf{u} = \mathbf{f}, \tag{2}$$

where the coefficients of $\mathbf{A}$ depend on the spatial samplings $(\Delta x, \Delta z)$, velocity $(\mathbf{c})$ and the frequency (ω). Note that $\mathbf{A}$ is nonlinear in the velocity $c(\mathbf{x})$ and is independent for each frequency.

EXAMPLES

The Helmholtz solver contains four principal operators. These operators consist of three wave equation operators: the nonlinear modeling operator, the linearized modeling operator, the adjoint of the linearized modeling operator and a receiver restriction operator that interpolates the wavefield onto the receiver locations. Each of the wave-equation operators has the capability of performing conventional phase and amplitude, phase-only and normalized-amplitude modeling. In the following subsections, we provide examples of how to use these operators on simple synthetic models.

Nonlinear forward operator

The nonlinear forward wave equation operator maps an input source function $\mathbf{f}$ to a data vector $\mathbf{d}$ given a velocity model:

$$\mathbf{d} = \mathbf{R}\mathbf{A}^{-1}\mathbf{f}, \tag{3}$$

where $\mathbf{A}$ is the Helmholtz matrix as defined in equation 2 and $\mathbf{R}$ is a restriction operator that interpolates the wavefield to the receivers. Note that intermediately, we compute the wavefield $\mathbf{u} = \mathbf{A}^{-1}\mathbf{f}$. When we do this, we only need to factorize $\mathbf{A}$ once and thus we can apply $\mathbf{A}^{-1}$ to multiple sources.

As $\mathbf{A}$ is quite large, sparse and asymmetric, sparse LU decomposition algorithms can be used to factorize $\mathbf{A}$ into two lower and upper triangular matrices. In this Helmholtz solver, this is done with the UMFPACK libraries within SuiteSparse.

To factorize $\mathbf{A}$ and perform the nonlinear modeling, we must construct both $\mathbf{A}$ and $\mathbf{f}$. To construct $\mathbf{A}$ we need to choose a velocity model and frequency. The velocity model for this particular example is shown in Figure 1 and the frequency of choice is 1.0 Hz. For all of these examples, the velocity contains no imaginary component. Note that another advantage of frequency domain modeling is that it is trivial to implement attenuation by providing an imaginary component to the velocity.

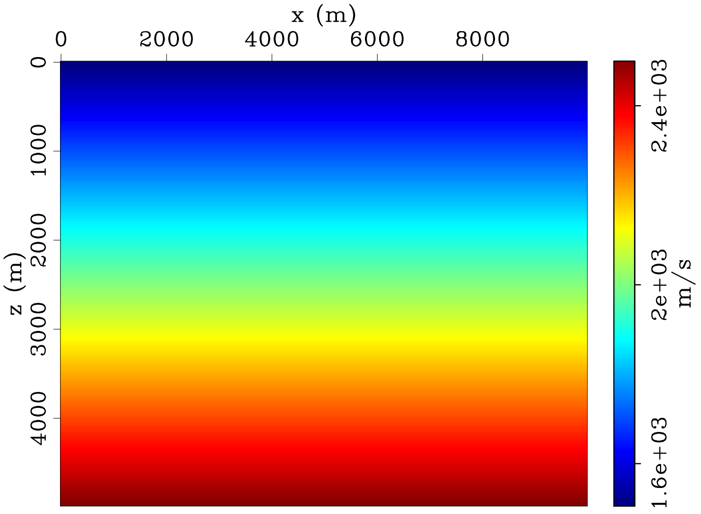


Figure 1: The velocity used in the nonlinear modeling. For this example it contains no imaginary component. [**ER**] joseph2/. Voz-vel

For velocity model shown in Figure 1 and a frequency of 1.0 Hz, the Helmholtz matrix $\mathbf{A}$ is shown in Figure 2. Note that, $\mathbf{A}$ is banded but not symmetric. This asymmetry is due to the use of one-sided derivatives used in the Engquist/Majda absorbing boundaries. (Ajo-Franklin, 2005; Engquist and Majda, 1977). The width of the bands depends on the length of the stencil used for computing the Laplacian. For higher frequencies and faster velocities, the stencil size must be increased to avoid numerical dispersion. This in turn increases the computational cost of factorizing $\mathbf{A}$. For this reason this solver implements a compact 9-point stencil that is based on a Padé-like approximation (Singer and Turkel, 1998).

In order to construct the source vector $\mathbf{f}$ and the receiver restriction operator $\mathbf{R}$, the solver must be provided with the source and receiver coordinates. Note that this solver does not assume that the sources and the receivers are located on the same grid as the velocity. Figure 3 shows the overlay of the acquisition geometry for the nonlinear modeling and also that the sources and receivers are placed on a different grid than the velocity grid.

With each of the components of equation 3 built, we can perform the nonlinear modeling. The real part of the result of the modeling performed is shown in Figure 4(a) and the real part of the recorded data is shown in Figure 4(b).

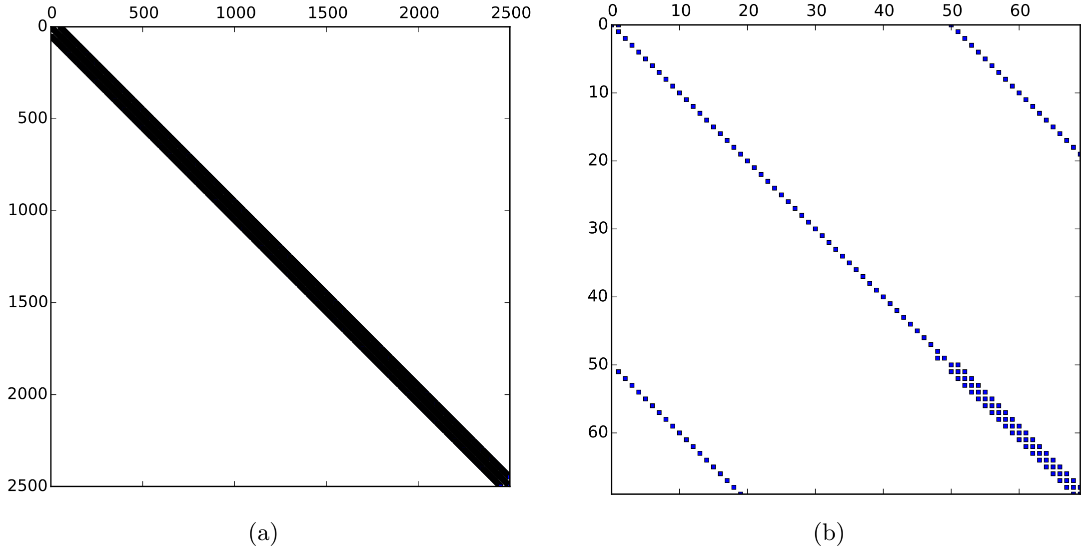


Figure 2: (a) The sparsity pattern of **A**. (b) A zoom-in on the upper diagonal of the matrix (first 70 elements). Note the three bands on the main diagonal and the outer fringes. Also note two stray points near the main diagonal. These stray points appear due to the one-sided derivative at the boundary and are the cause of the asymmetry of **A**. [**ER**] joseph2/. A-0-2499,A-0-69

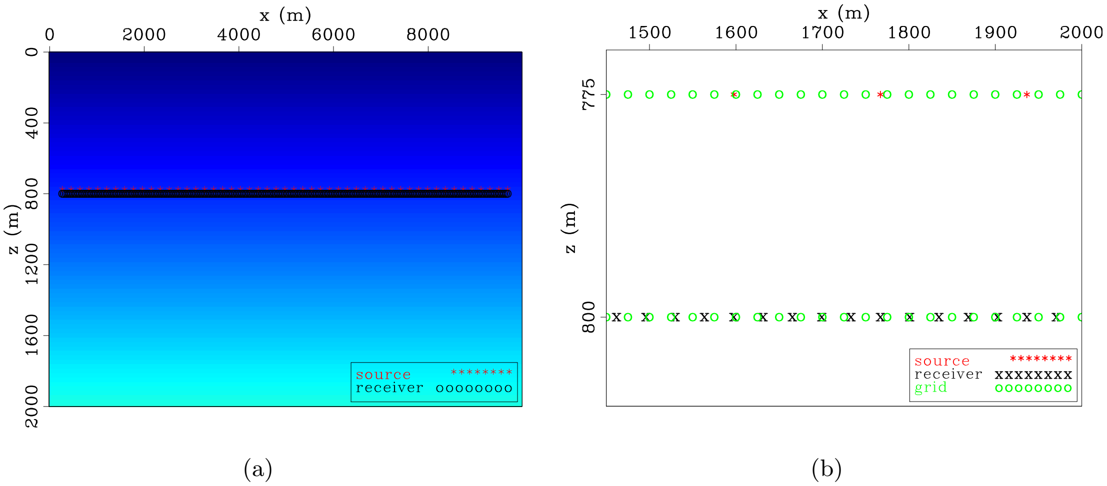
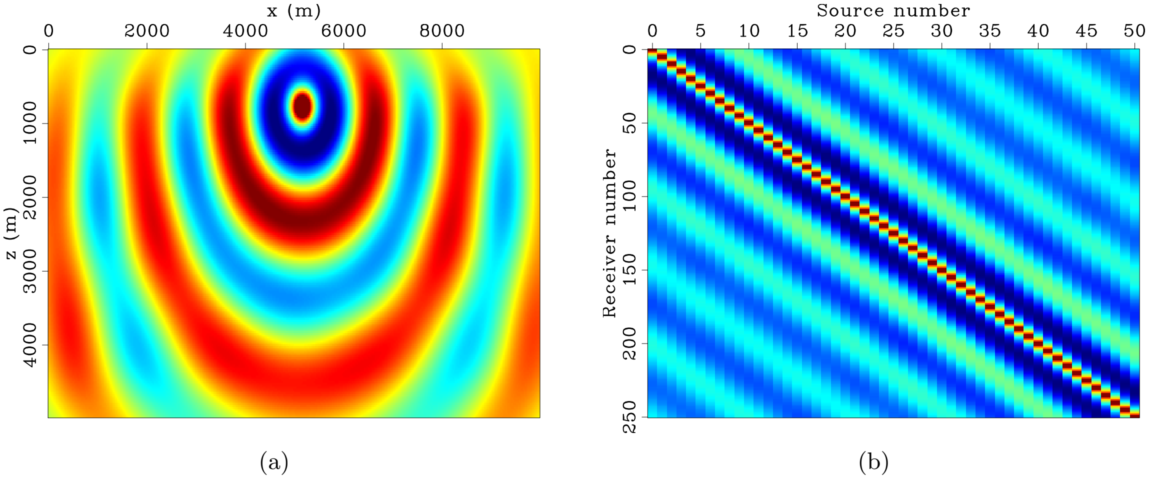


Figure 3: (a) The acquisition geometry for the nonlinear modeling overlain on the velocity model shown previously in Figure 1. Note that the velocity here is shown only to 2000 m in depth. (b) A zoom-in on the source, receiver and velocity grid. Note that the source, receiver and velocity grid all have different spatial sampling intervals and are not co-located. [**ER**] joseph2/. vel-geom-ovrly,srsrecmod



(a) (b)

Figure 4: (a) The real part of the wavefield for a wavelet of 1Hz .(b) The real part of the data obtained after modeling at all source locations and the application of **R**. [**ER**] joseph2/. wfldreal,nlreal

Linearized operator

We can express the linearized operator in terms of the Helmholtz matrix by first expressing $\mathbf{A}^{-1}$ as a first-order Taylor-series expansion around $\mathbf{m_0}$:

$$\mathbf{A}^{-1}(\mathbf{m})\mathbf{f} \approx \mathbf{A}^{-1}(\mathbf{m_0})\mathbf{f} + \left(\frac{\partial \mathbf{A}^{-1}(\mathbf{m_0})}{\partial \mathbf{m}}\Delta\mathbf{m}\right)\mathbf{f},$$

where $\mathbf{m}$ is a vector of length nm containing the velocity $c(\mathbf{x})$ and $\mathbf{m} = \mathbf{m_0} + \Delta\mathbf{m}$. $\mathbf{m_0}$ is the initial (background) model and $\Delta\mathbf{m}$ is our model perturbation. We can further expand the derivative as:

$$\Delta\mathbf{u} = \left(\frac{\partial \mathbf{A}^{-1}(\mathbf{m_0})}{\partial \mathbf{m}}\right)\mathbf{f} = \left(-\mathbf{A}^{-1}\frac{\partial \mathbf{A}(\mathbf{m_0})}{\partial \mathbf{m}}\mathbf{A}^{-1}\Delta\mathbf{m}\right)\mathbf{f}, \tag{4}$$

where $\Delta\mathbf{u}$ is the linearized Born wavefield. As $\partial\mathbf{A}/\partial\mathbf{m}$ is a tensor, we need to sum over the different components of $\mathbf{m}$ to write equation 4 as a chain of operators:

$$\Delta\mathbf{u} = -\mathbf{A}^{-1}(\mathbf{m_0})\sum_{i}^{nm}\frac{\partial \mathbf{A}(\mathbf{m_0})}{\partial m_i}\Delta m_i\mathbf{A}^{-1}(\mathbf{m_0})\mathbf{f}. \tag{5}$$

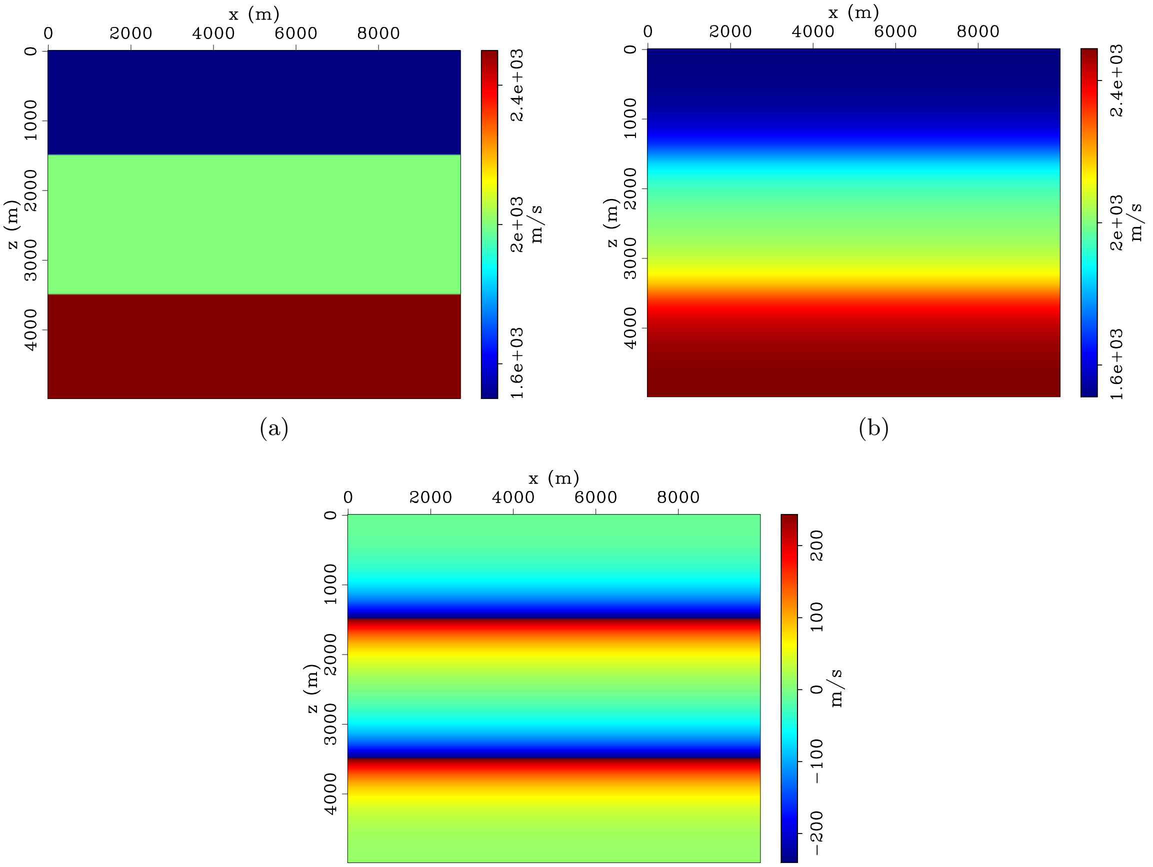
At this point in our derivation, the physical interpretation is quite clear. Starting from the rightmost side of equation 5, we have the background wavefield from the application of $\mathbf{A}^{-1}$ to the source $\mathbf{f}$. We denote this as $\mathbf{u_0}$. This is then scaled by a perturbation and a second time derivative is applied with the application of $\frac{\partial \mathbf{A}}{\partial m_i}$. The output of these two operations ($\frac{\partial \mathbf{A}}{\partial m_i}\mathbf{u_0}\Delta m_i$) is commonly known as the secondary source in linearized modeling, and serves as input to another propagation $\mathbf{A}^{-1}$, thus resulting in the single-scattered or Born wavefield ($\Delta\mathbf{u}$). If we consider only perturbing one single point in the model space (i.e. $\Delta\mathbf{m}$ contains only one non-zero component), at model point j then we can drop the summation and rewrite equation 5 as:

$$\Delta\mathbf{u} = \mathbf{A}^{-1}\mathbf{D}\mathbf{u_0}\Delta m,$$

where $\mathbf{D} = \frac{\partial \mathbf{A}}{\partial m_j}$ and $\Delta m = \Delta m_j$. With the application of $\mathbf{R}$, we now have a linear relationship between a perturbation in the data and a perturbation in the model parameters:

$$\Delta\mathbf{d} = \mathbf{R}\Delta\mathbf{u} = \mathbf{R}\mathbf{A}^{-1}\mathbf{D}\mathbf{u_0}\Delta m. \tag{6}$$

To perform the linearized modeling in the solver, we use the layered velocity which consists of the smooth background and the perturbation. These are shown in Figure 5. The real part of the wavefield that resulted from applying the chain of operators in equation 5 to the perturbation shown in Figure 5(c) is shown in Figure 6(a). Figure 6(b) shows the recorded data.

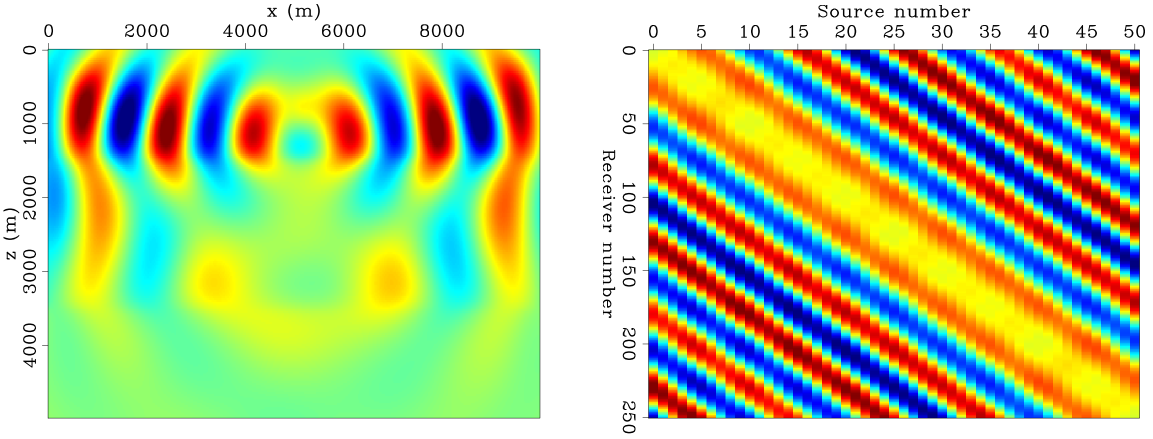


(a)

(b)

(c)

Figure 5: (a) Total velocity model ($\mathbf{m} = \mathbf{m_0} + \Delta\mathbf{m}$), (b) background velocity $\mathbf{m_0}$ (low-pass filtered version of (a)) and (c) velocity perturbation $\Delta\mathbf{m}$. [**ER**] joseph2/. Layer-vel,Smooth-vel,Pert-vel



(a) (b)

Figure 6: (a) The real part of the Born wavefield modeled $\mathbf{\Delta u}$ as result of applying the chain of operations in equation 5 to to the perturbation shown in Figure 5(c). (b) The real part of the data perturbation obtained after Born modeling at all source locations and the application of **R**. [**ER**] joseph2/. lwfldreal,llyr-real

Adjoint of linearized operator

With a simple conjugation and reversal of the operators in equation 6 we obtain the expression for the application of the adjoint Born operator that relates a perturbation in the data to a perturbation in the model parameters:

$$\Delta m = \mathbf{u_0}^* \mathbf{D}^* (\mathbf{A}^{-1})^* \mathbf{R}^* \Delta \mathbf{d}.$$

Using the solver, we first attempt to map the data perturbation computed in Figure 6(b) back to the model space. This is shown in Figure 7(a). In this migrated image, we can faintly see the first velocity contrast that occurs at 1500 m depth (Figure 5(c) is included for reference). As a second example, we compute a monochromatic sensitivity kernel for homogeneous velocity that is shown in Figure 8.

CONCLUSION

We have presented a Helmholtz solver written in C++ that can be used for performing frequency domain FWI. This solver contains three main operators that implicitly solve the Helmholtz equation to compute the monochromatic non-linear, linearized and adjoint wavefields and data. While in this report only conventional amplitude plus phase modeling was performed, this solver has the machinery to also perform phase-only and normalized-amplitude modeling as well. While factorizing the Helmholtz matrix is more involved than solving wave equation via explicit finite differences, the frequency domain implementation has many advantages such as frequency-continuation, straight-forward implementation of attenuation and cheap multi-source modeling. We have presented examples of performing forward modeling and computing gradients in hopes that this software may be used among SEP researchers.

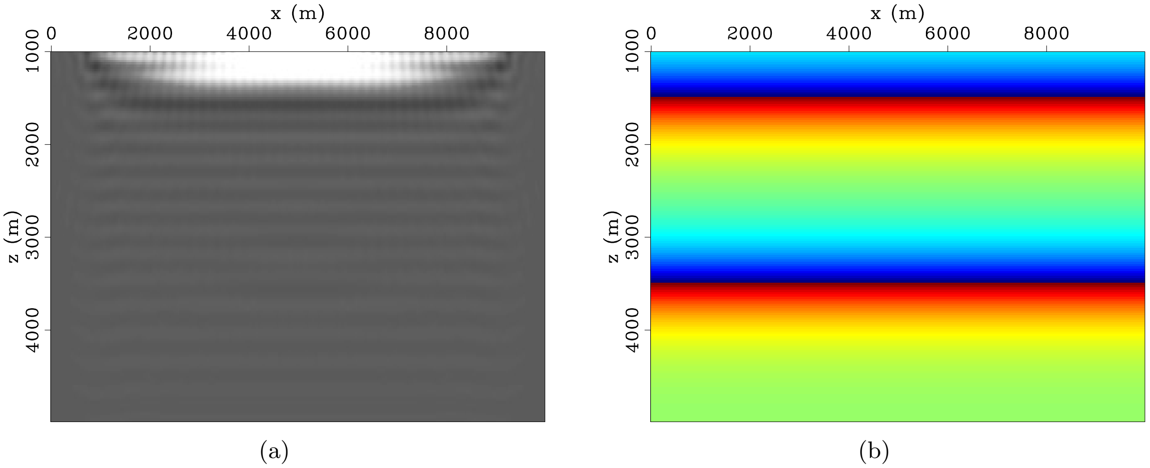


Figure 7: (a) The adjoint of the linearized operator applied to the data perturbation shown in Figure 6(b). (b) The model perturbation in Figure 5(c) shown again for comparison. Note that the image and perturbation are shown from 1000 m depth. Also note the faint reflector at 1500 m depth in the image that corresponds to the first reflector in the model. [**ER**] joseph2/. lyrimg-4,Pert-velcut

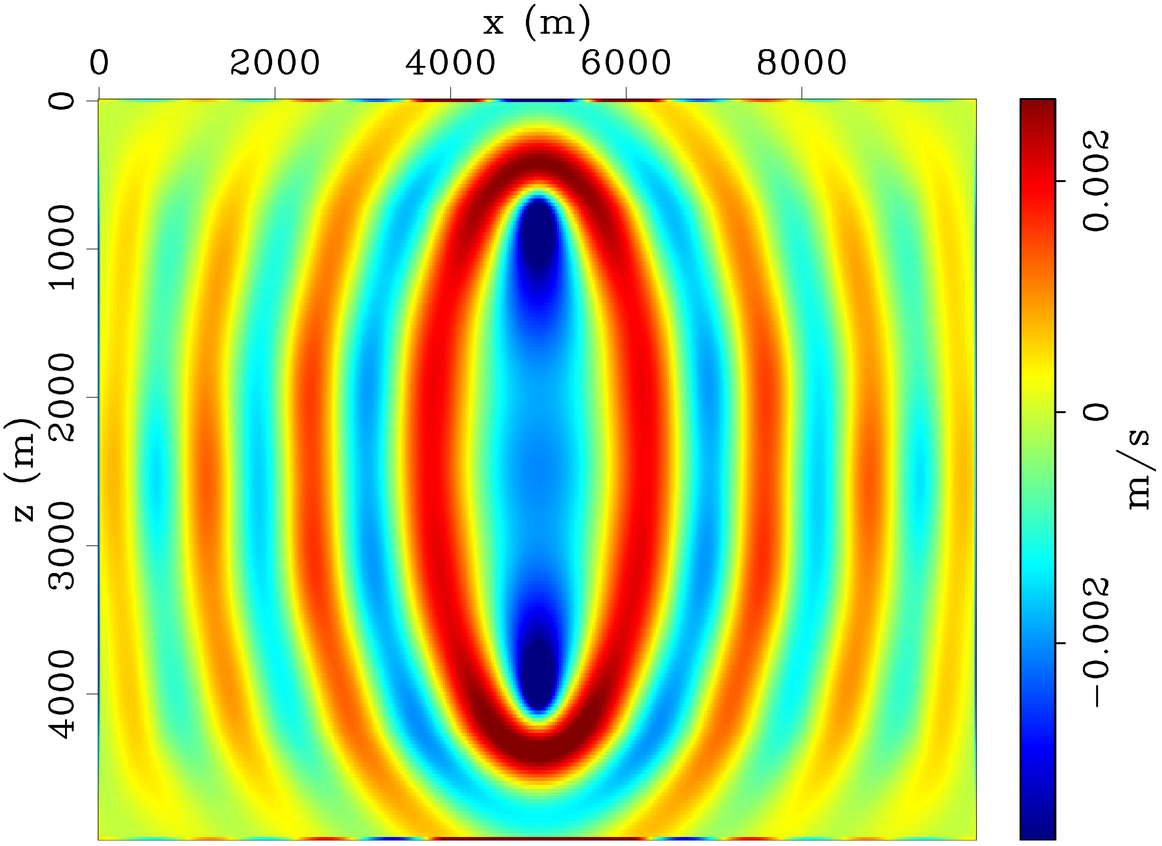


Figure 8: FWI gradient (sensitivity kernel). Note that this example used only one source and one receiver located at 775 m and 4000 m depth respectively. [**ER**] joseph2/. sens-kern

ACKNOWLEDGEMENT

The first author would like to thank the sponsors of the SEP for financial and intellectual support. He would also like to thank Rahul Sarkar and Huy Le for their help in deriving the linearized operator.

REFERENCES

Ajo-Franklin, J. B., 2005, Frequency-domain modeling techniques for the scalar wave equation: An introduction.

Biondi, B. and A. Almomin, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79**, WA129–WA140.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Davis, T., W. Hager, and I. Duff, 2014, Suitesparse: htt p://faculty. cse. tamu. edu/davis/-suitesparse. html.

Engquist, B. and A. Majda, 1977, Absorbing boundary conditions for numerical simulation of waves: Proceedings of the National Academy of Sciences, **74**, 1765–1766.

Marfurt, K. J., 1984, Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations: Geophysics, **49**, 533–549.

Mora, P., 1987, Elastic wavefield inversion: PhD thesis, Stanford University.

Pratt, R. G., C. Shin, and G. Hick, 1998, Gauss–newton and full newton methods in frequency–space seismic waveform inversion: Geophysical Journal International, **133**, 341–362.

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

Singer, I. and E. Turkel, 1998, High-order finite difference methods for the helmholtz equation: Computer Methods in Applied Mechanics and Engineering, **163**, 343–358.

Wang, C., J. Park, R. Bloor, J. Leveille, and P. Farmer, 2011, Application of time domain and single frequency waveform inversion to real data: Presented at the 73rd EAGE Conference and Exhibition incorporating SPE EUROPEC 2011.

Washbourne, J., N. Shah, and K. Bube, 2013, System and method for data inversion with phase extrapolation. (WO Patent App. PCT/US2012/039,077).

Stanford Exploration Project, SEP165, October 8, 2016

Reproducibility through containers

Robert G. Clapp

ABSTRACT

In order to achieve truly reproducible results the underlying software architecture must be captured. Docker containers, a lightweight alternative to a virtual machine, can be used to snapshot all software dependencies and allow anyone to reproduce an author's results with minimal effort. We demonstrate the effectiveness of Docker containers in several contexts including reproducible research, computer labs, and writing LaTeX documents.

INTRODUCTION

Reproducible research is one of the fundamental building blocks of scientific advancement. Claerbout (1990) and Schwab et al. (1996) designed a framework for authors to follow to make their work reproducible. Their basic concept was to introduce generic targets of `build` (build all results), `view` (view the results), `burn` (remove the results), `clean` (remove all intermediate files) that each author would define for each paper. A limitation of their approach is that when their underlying software dependencies changed, reproducibility (or even the ability to compile) is not guaranteed. Fomel and Claerbout (2009) went a step further incorporating papers into the software building process. As a result, in theory, papers stay reproducible because any bugs introduced into the software are caught and fixed in the build process. This approach, in addition to producing an ever larger, and more complex building/debugging requirement, still makes assumptions that all possible underlying software requirements have been tested/debugged.

The difference in development and production environments is a well known problem in the broader software development field. Virtualization, which creates a virtual machine running within a system, is one approach. Virtualization has drawbacks such as slow provisioning, performance degradation, and a large memory footprint when running multiple virtual machines on a single server. A newer approach is containers (Wikipedia, 2016). The current leading container approach are 'Docker containers'. In the context of reproducible research, a Docker container will include: the rules to build a paper, the software needed to build the paper, the underlying operating system which the user built the paper on, and how that underlying operating system was constructed.

In this paper we demonstrate how to use Docker containers to build a reproducible environment. We demonstrate their utility by using them for report articles, labs, and even building documents.

INTRODUCTION TO DOCKER CONTAINERS

At a very basic level a container can be thought of as a virtual machine. You have an additional operating system that is taking a part of a host's resources. In general, the operating system of the virtual machine exists in the computer's memory instead of on disk. The advantage of the virtual machine concept is that we can construct the perfect environment for my application to run in. In terms of reproducible research, once we've tested that the code works in a given virtual machine, that working state is preserved forever. We can distribute the virtual machine image to anyone and it will simply work without much, or any, effort on their part. In addition, we can run multiple virtual machine images on a single host, each running completely independently.

The complete independence of each virtual machine is also one of its major drawbacks. Imagine using virtual machines to do a parallel task, because each virtual machine is completely independent each will have its own complete copy of the operating system. Containers work a little differently. We normally think of a filesystem as consisting of a series of directories and files. The directories and files may sit on different disks or servers but only a single version of a given directory exists for a given machine. As we make a Docker image, we are building up the filesytem in a series of layers. Each command in the Docker build process takes a difference between the state of the filesytem before and after a build command. Each layer of the Docker image is read-only. By default any changes we make to the Docker filesytem while running a given Docker image is making changes to a temporary additional filesystem layer. When we exit, all of these changes will be lost (we will discuss read/write filesystem layers later). When running multiple Docker virtual machines the read-only layers will be shared between the Docker instances resulting in a much smaller memory footprint.

In addition Dockers run on the host operating system, allowing it to share a lot of the host resources. As a result while a virtual machine can take minutes to start a Docker often starts in less than a second.

Building a Docker image

A Docker image is built from a `Dockerfile`. A `Dockerfile` consists of a series of commands to build the Docker image. As an example we are going to step through a `Dockerfile` that builds SEPlib on a Centos-7 machine.

```
From centos:7
MAINTAINER Bob Clapp <bob@sep.stanford.edu>
```

The beginning of my `Dockerfile` indicates that we want to start with another Docker image, in this case, CentOS version 7 (the colon is how you indicate a version number to Docker). SEPlib uses SU's library for reading SEG-Y headers so before we build SEPlib we are going to build SU. The CentOS image is quite small, with the minimum number of packages. In order to build SU we need to install sum X11 libraries, make, gcc, and wget using yum. The `RUN` keyword indicates system commands that need to be run to build the image.

```
RUN yum -y install make libX11-devel libXt-devel gcc wget; yum -y clean all
```

After we've installed these packages we need to create a directory for SU and then download and compile it. We need to do a little bit of work because SU wants to have this section be interactive.

```
RUN mkdir /opt/SU && cd /opt/SU ; \
   wget ftp://ftp.cwp.mines.edu/pub/cwpcodes/cwp_su_all_44R1.tgz  &&\
  cd /opt/SU ; tar xf cwp_su_all_44R1.tgz;cd /opt/SU/src   &&\
  cd /opt/SU/src ;touch cwp_su_version LICENSE_44_ACCEPTED MAILHOME_44 &&\
  cd /opt/SU/src;  echo "echo boo"  >chkroot.sh &&\
  cd /opt/SU/src; chmod 777 chkroot.sh   &&\
  cd /opt/SU/src; CWPROOT=/opt/SU make install xtinstall &&\
   rm -rf /opt/SU/cwp_su_all_44R1.tgz /opt/SU/bin
```

Docker compares the filesystem before and after each command. By combining all of the installation steps and cleanup on a single line we reduce the overall size of my Docker image. Once we have the parts of the SU that we need to install some additional software packages the SEPlib uses that SU does not.

```
RUN yum -y install make automake autoconf libtool csh git \
 libXt-devel libX11-devel libXaw-devel  gcc gcc-gfortran flex &&
  yum -y clean all
```

We need to download, compile, and install SEPlib.

```
RUN mkdir /opt/SEP &&\
    git clone http://zapad.Stanford.EDU/bob/SEPlib.git /opt/SEP/src && \
    cd /opt/SEP/src; autoreconf -vif  &&\
    cd /opt/SEP/src; ./configure --prefix=/opt/SEP  --with-su=/opt/SU && \
    cd /opt/SEP/src;  make install &&\
    cd /opt/SEP/src;  make clean
```

Finally, we are going to add the environmental variables SEPlib needs to the root user's environment.

```
RUN echo export PATH=$PATH:/opt/SEP/bin >> ~/.bash_profile &&\
    echo export SEP=/opt/SEP >> ~/.bash_profile &&\
    echo export SEPINC=/opt/SEP/include >> ~/.bash_profile &&\
    echo export PYTHONPATH=/opt/SEP >> ~/.bash_profile
```

Once we've written my `Dockerfile` we can build my Docker image using

```
 docker build -t rgc007/seplib:8-centos .
```

where `-t rgc007/seplib:8-centos` indicates that the tag `-t` for this image is `seplib` version `8-centos` for the Docker account `rgc007`. The `.` indicates the directory where the `Dockerfile` and any additional files we might add to my image exist. After the image is built we can run:

```
docker run -it rgc007/seplib:8-centos /bin/bash
```

The run command will check to see if the image exists locally. If not, it will it will attempt to download the layers needed for the Docker image from from the Docker account `rgc007`. Any layers that don't exist locally will then be downloaded. Finally it will give me a bash shell within the container containing a full version of SEPlib. we can push this image to Docker by using the command `docker push rgc007/seplib:8-centos`. Once Docker is installed the reader can reproduce the SEPlib build by typing `make buildSEPRepo` or enter the Docker image by typing `make enterSEP`.

USING DOCKERS

There are many uses for containers. Below we discuss four examples of using Dockers in a research university environment. The Dockerfile described above sets up an image with a basic SEPlib environment. My reproducible research Docker is going to begin by inheriting my SEPlib image.

```
From rgc007/seplib:8-centos
MAINTAINER Bob Clapp <bob@sep.stanford.edu>
```

We could follow the same procedure described above to build and enter my Docker but we wouldn't be able display any X11 graphics. There are several different options to get graphics working. The one we are going to choose is to add a local ssh server to my Docker. Using a ssh server approach has the advantage of working on Linux, MacOSX, and with the right ports and an X11 client, Windows. First we need to install ssh, passwd, and xauth.

```
RUN yum -y install  openssh-server passwd  xauth; yum clean all
```

Next we are going to add to the Docker a script that creates a non-root user, sets its password, and copy roots environment. In addition it is going to set the SEPlib datapath to
`tmp`.

```
ADD start.sh
```

where start.sh contains

```
#!/bin/bash

__create_user() {
# Create a user to SSH into as.
useradd user
SSH_USERPASS=newpass
echo -e "$SSH_USERPASS\n$SSH_USERPASS" | (passwd --stdin user)
echo ssh user password: $SSH_USERPASS
echo datapath=/tmp/ >~user/.datapath
```

```
cp ~root/.bash_profile ~user/.bash_profile
}

# Call all functions
__create_user
```

To run the ssh daemon we need to create an additional directory and generate an ssh key.

```
RUN mkdir /var/run/sshd
RUN ssh-keygen -t rsa -f /etc/ssh/ssh_host_rsa_key -N ''
```

We are then going to run the `start.sh` script.

```
RUN chmod 755 /start.sh
RUN ./start.sh
```

We are then going to add to my new Docker a local folder `derivative` containing a Makefile and a simple code that applies a first derivate to a 2-D field.

```
ADD derivative /home/user
RUN chown -R user /home/user
```

Finally we need to indicate that to enter this image that to enter it one should use ssh.

```
ENTRYPOINT ["/usr/sbin/sshd", "-D"]
```

Once we've built this image (e.g. `docker build -t rgc007/testReport`), We can start the image using

```
docker run -d -p 22  -t rgc007/testReport
```

in this case we are running the Docker in detached mode and mapping port 22 of the Docker to an available port on my local machine. We can get what port my Docker image mapped port 22 to using `docker ps`. Finally, we can login to my Docker image using `ssh -Y user@localhost -p XXXX` where XXXX is the port number I found from the `docker ps` command[1] We use the password `newpass` to login and We are now in a complete linux environment where we can use make `build, burn, clean, view`. The reader can build the reproducibility Docker by typing `make buildSEPRepo`. You will need to enter the password `newpass` at the prompt.

Another use, in an academic research context, is to use a Docker to build LaTex environment. The Docker image consists of a series of read-only file system layers. As a result even though we can modify a file within the Docker filesystem those changes will be lost once the Docker image is stopped. For the reproducible research example above the

---

[1] It is useful to add -Y -o UserKnownHostsFile=/dev/null -o StrictHostKeyChecking=no to avoid having to edit your known host file.

read-only nature of the file system is sufficient. When writing a paper we need to have some persistency in our filesystem. The solution to what Docker refers to as Docker data volume. The basic concept is to map a directory from our host system into a directory inside Docker images. Any changes made in the directory of the Docker image is seen on the host and vice versa. To create a Docker volume we simply add an option to the `docker run` command line option of the form `--volume=/local/path:/docker/path` where `local/path` refers to a directory on the host machine and `/docker/path` refers to the directory in Docker image. The reader can build the SEP LaTeX environment by typing `make buildSEPTeX` and enter the image by typing `make enterSEPTeX`.

SEP has a series of four classes that it requires every student to take. One of the challenges are making sure that peculiarities of a given SEPlib release, compiler, user environment, etc. do not distract from the purpose of the lab. Dockers solve this problem. Once a Docker image is built and works once it will work without needed additional changes. This also allows us to provide our labs to the world without the challenges of completely setting up their environment. The lab Docker combines elements of the LaTeX and reproducible research Docker. The Docker can be built using `make buildSEPlLab` and entered using `make enterSEPLab`, again using the `newpass` password.

Another interesting use for a Docker is to distribute single software executables. The entrypoint of a Docker can be running a unix command. When used in conjunction with a Docker data volume, theoretically a full processing system could be written using Docker containers. A `CMD` is added the `Dockerfile` specifying what program should be run and any potential arguments. As an example typing `make showModel` will first build a 3-D synthetic model and then display it using SEPlib's 3-D viewer.

CONCLUSIONS

Dockers represent a more complete step along the path of reproducible research. They can be used to capture not only the user's code but the user's full environment. Dockers also provide effective solutions for code distribution, computational labs, and building LaTeX documents.

REFERENCES

Claerbout, J. F., 1990, Active documents and reproducible results: SEP-Report, **67**, 139–144.

Fomel, S. and J. F. Claerbout, 2009, Reproducible research: Computing in Science & Engineering, **11**, 0005–7.

Schwab, M., M. Karrenbach, and J. Claerbout, 1996, Making scientific computations reproducible: SEP-Report, **92**, 327–342.

Wikipedia, 2016, Operating-system-level virtualization — Wikipedia, the free encyclopedia. ([Online; accessed 23-September-2016]).

SEP PHONE DIRECTORY

Name	Phone	Login Name
Alves, Gustavo	723-9282	gcalves
Barak, Ohad	723-9282	ohad
Barnier, Guillaume	723-6006	gbarnier
Baroni, Claudia	723-5002	
Biondi, Biondo	723-1319	biondo
Biondi, Ettore	723-9282	ettore
Cabrales, Alejandro	723-6006	cabrales
Chang, Jason	724-4322	jasonpc
Claerbout, Jon	723-3717	jon
Clapp, Bob	725-1334	bob
Dahlke, Taylor	724-4322	taylor
Farris, Stuart		stuart
Huot, Fantine	723-0463	fantine
Jennings, Joe	723-1250	joseph29
Le, Huy	723-1250	huyle
Levin, Stewart	726-1959	stew
Ma, Yinbin	723-0463	yinbin
Martin, Eileen	723-0463	ermartin
Sarkar, Rahul	725-1625	rahul
Shen, Yi	723-6006	yishen

Dept fax number: (650) 725-7344

E-MAIL

Our Internet address is "*sep.stanford.edu*"; i.e., send Jon electronic mail with the address "*jon@sep.stanford.edu*".

WORLD-WIDE WEB SERVER INFORMATION

Sponsors who have provided us with their domain names are not prompted for a password when they access from work. If you are a sponsor, and would like to access our restricted area away from work, visit our website and attempt to download the material. You will then fill out a form, and we will send the username/password to your e-mail address at a sponsor company.

STEERING COMMITTEE MEMBERS, 2016-2017

Name	Company	Tel #	E-Mail
Raymond Abma	BP	(281)366-4604	abmar1@bp.com
Francois Audebert	TOTAL	(33)-6-1278-4195	francois.audebert@total.com
Biondo Biondi	SEP	(650)723-1319	biondo@sep.stanford.edu
Jon Claerbout	SEP	(650)723-3717	jon@sep.stanford.edu
Thomas Dickens	ExxonMobil	(713)431-6011	tom.a.dickens@exxonmobil.com
Carlo Fortini (Co-chair, 1st year)	ENI	+39 02-520-61599	carlo.fortini@eni.com
Faqi Liu	Hess	(823)231-9218	fliu@hess.com
Yi Luo	Saudi Aramco	–	yi.luo@aramco.com
Guojian Shan (Co-chair, 1st year)	Chevron	–	guos@chevron.com
Alejandro Valenciano	PGS	–	alejandro.valenciano@pgs.com
Kathy Zou	PGS	–	kathy.zou@pgs.com

Research Personnel

Ohad Barak received a B.Sc. (2006) and an M.Sc. (2009) in Geophysics from Tel-Aviv University. In 2008 he joined the R&D team of Paradigm Geophysical, maintaining and developing the production codes. He joined SEP in 2009 and is currently pursuing a Ph.D. in geophysics at Stanford University, and a longer biography. His current research involves seismic rotation data and the various methods of acquiring them, on land and on the ocean bottom. He is working on showing how rotation data can be utilized in combination with standard geophone and hydrophone data for applications such as wave-mode separation and data reconstruction.

Biondo L. Biondi is professor of Geophysics at Stanford University. Biondo graduated from Politecnico di Milano in 1984 and received an M.S. (1988) and a Ph.D. (1990) in geophysics from Stanford. He is co-director of the Stanford Exploration Project and of the Stanford Center for Computational Earth and Environmental Science. In 2004 the Society of Exploration Geophysicists (SEG) has honored Biondo with the Reginald Fessenden Award. Biondo recently published a book, 3-D Seismic Imaging, that is the first text book to introduce the theory of seismic imaging from the 3-D perspective. The book is published by SEG in the Investigations in Geophysics series. During 2007 gave a one-day short course in 28 cities around the world as the SEG/EAGE Distinguished Short Course Instructor (DISC) . He is a member of AGU, EAGE, SEG and SIAM.

Alejandro Cabrales-Vargas obtained his Bachelors degree in Geophysics in the University of Mexico in 2002. He has being working for Petroleos Mexicanos since 2002, initially in seismic interpretation for oil and gas exploration, and more recently in the supervision of depth imaging processes. He obtained his Masters Degree in Geophysics in the University of Oklahoma in 2011. He joined SEP in the fall of 2014, and is currently working towards his PhD.

Robert Clapp received his B.Sc. (Hons.) in Geophysical Engineering from Colorado School of Mines in May 1993. He joined SEP in September 1993, received his Masters in June 1995, and his Ph.D. in December 2000. He is a member of the SEG and AGU.

Stuart Farris received his B.S. in Geophysical Engineering from the Colorado School of Mines with an added minor in Computer Science. After graduating in May of 2016, he joined the Stanford Exploration Project that summer and began his Ph.D. at Stanford University. He is interested in the unique problems associated with processing land seismic data and tomographic full waveform inversion.

Fantine Huot received her M.S. in Science and Engineering from the Ecole Normale Supérieure des Mines de Paris in 2013. Following graduation, she worked as a software developer in France. In fall 2015, she joined the SEP and started her Ph.D. at Stanford University. She is interested in optimal acquisition design and machine learning.

Joseph Jennings Joseph received his BS in Geophysical Engineering from Colorado School of Mines in 2014. Following graduation, he worked in land seismic processing R&D at ION Geophysical. In the summer of 2015, he joined the SEP as a PhD student where he hopes to pursue research in simultaneous source imaging and imaging of unconventional reservoirs.

Huy Le earned his B.S. degree in Geophysics from the University of Oklahoma in 2012 and his M.S. degree in Computational Geoscience from Stanford University in 2014. He is now a PhD candidate in Geophysics. His research interests include anisotropy, full waveform inversion, finite difference method, and high-performance computing.

Eileen Martin graduated from the University of Texas at Austin in 2012 with a B.S. in Mathematics and Computational Physics. She is a Department of Energy Computational Science Graduate Fellow, and is working towards her Ph.D. in the Institute for Computational and Mathematical Engineering. She Joined SEP in the summer of 2013.

Shuki Ronen is currently (2015) external and collaborative research manager of Dolphin Geophysical and a consulting faculty at SEP.

Rahul Sarkar graduated with a B.S and M.S from the Indian Institute of Technology, Kharagpur in 2011. After that he worked for Schlumberger for four years in the United States and Mexico. Since 2015, he is pursuing a M.S in the Institute for Computational and Mathematical Engineering at Stanford University.

Yi Shen got her Bachelors degree in Acoustics, from Dept. of Electronic Science and Engineering, Nanjing University, China in 2010. She joined SEP in the fall of 2010, and is currently working toward a Ph.D. in Geophysics at Stanford University. She is a student member of the SEG.

John Washbourne is a Senior Staff Geophysicist at Chevron. He received his Ph.D. in Engineering Geoscience from the University of California, Berkeley in 1998.

Ethan Williams is a fourth year undergraduate student in the Department of Geophysics. He worked as an intern with SEP in the summer of 2016, and plans to pursue a Ph.D. in seismology after graduation.

SEP ARTICLES PUBLISHED OR IN PRESS

Almomin, A., and Biondi, B., 2012, Tomographic full waveform inversion: Practical and computationally feasible approach. SEG Technical Program Expanded Abstracts, 1–5.

Almomin, A., and Biondi, B., 2013, Tomographic full waveform inversion (TFWI) by successive linearizations and scale separations. SEG Technical Program Expanded Abstracts, 1048–1052.

Almomin, A., and Biondi, B., 2014, Preconditioned tomographic full waveform inversion by wavelength continuation. SEG Technical Program Expanded Abstracts, 944–948.

Almomin, A., and Biondi, B., 2014, Generalization and amplitude normalization of tomographic full waveform inversion. SEG Technical Program Expanded Abstracts.

Barak, O., Jaiswal, P., de Ridder, S., Giles, J., Brune, R. and Ronen, S., 2014, Six-component seismic land data acquired with geophones and rotation sensors: wave-mode separation using 6C SVD: SEG Technical Program Expanded Abstracts (Submitted).

Chang, J., de Ridder, S., and Biondi, B., 2013, Power spectral densities and ambient noise cross-correlations at Long Beach: SEG Technical Program Expanded Abstracts **32**, 2196–2200.

Chang, J., Nakata, N., Clapp, R.G., de Ridder, S., and Biondi, B., 2014, High-frequency surface and body waves from ambient noise cross-correlations at Long Beach, CA: SEG Technical Program Expanded Abstracts **33**.

Chang, J., and B. Biondi, 2015, Rayleigh-wave tomography using traffic noise at Long Beach, CA: SEG Technical Program Expanded Abstracts **34**.

Chang, J.P., de Ridder, S.A.L., and Biondi, B.L., 2016, High-frequency Rayleigh-wave tomography using traffic noise from Long Beach, California: Geophysics, 81, No. 2, B43–B53.

Dahlke, T., Beroza, G., Chang, J., and de Ridder, S., 2014, Stochastic variability of velocity estimates using eikonal tomography on the Long Beach data set: SEG Technical Program Expanded Abstracts **33** .

Dahlke, T., Biondi, B., Clapp, R., 2015, Domain decomposition of level set updates for salt segmentation: SEG Technical Program Expanded Abstracts **33** (Submitted).

Grobbe, N., F.C. Schoemaker, M.D. Schakel, S.A.L. de Ridder, E.C. Slob, and D.M.J. Smeulders, 2012, Electrokinetic fields and waves: Theory, experiments, and numerical modeling. Geophysical Research Abstracts, Vol. **14**, EGU2012–10636.

de Ridder, S., and B. Biondi, 2012, Continuous passive seismic monitoring of CCS projects by correlating seismic noise: A feasibility study. 74th Conference & Technical Exbition, EAGE, Extended Abstracts (Accepted).

de Ridder, S., and B. Biondi, 2012, Reservoir monitoring by passive seismic interferometry. SEG/SPE/AAPG Joint Workshop, La Jolla, CA, 24-29 June 2012 (Accepted).

de Ridder, S., 2012, Ambient seismic noise correlations for reservoir monitoring. SEG Technical Program Expanded Abstracts **31** (Submitted).

de Ridder, S., and B. Biondi, 2012, Continuous reservoir monitoring by ambient seismic noise tomography. SEG Technical Program Expanded Abstracts **31** (Submitted).

Schoemaker, F.C., N. Grobbe, M.D. Schakel, S.A.L. de Ridder, E.C. Slob, and D.M.J. Smeulders, 2012, Experimental validation of the electrokinetic theory and development of seismoelectric interferometry by cross-correlation: International Journal of Geophysics (Accepted).

Farghal, M.S., and Levin, S.A., 2012, Hunting for microseismic reflections using multiplets: SEG Technical Program Expanded Abstracts **31** (Submitted).

Guitton, A., Ayeni, G., and Esteban, D.A., 2012, Constrained full-waveform inversion by model reparameterization: Geophysics, 77, No. 2, R117–R127.

Guitton, A., and Esteban, D.A., 2012, Attenuating crosstalk noise with simultaneous source full waveform inversion: Geophysical Prospecting (In Press).

Guitton, A., 2012, Blocky regularization schemes for full waveform inversion: Geophysical Prospecting (In Press).

Halpert, A., 2012, Edge-preserving smoothing for segmentation of seismic images: SEG Technical Program Expanded Abstracts **31**, 1–5.

Halpert, A., Clapp, R.G., and B. Biondo, 2014, Salt delineation via interpreter-guided seismic image segmentation, Interpretation: **2**, T79–T88.

Arevalo, Humberto and Stewart A. Levin, 2014, Well and Seismic matching with ArcGIS and ProMAX via KML: Esri International User Conference Paper 901, San Diego, 15 July.

Leader, C., and R. Clapp, 2012, Least squares reverse time migration on GPUs - balacing IO and computation: 74th Conference & Technical Exbition, EAGE, Extended Abstracts (Accepted).

Leader, C., and A. Almomin, 2012, How incoherent can we be? Phase encoded linearised inversion with random boundaries: SEG Technical Program Expanded Abstract **31** (Submitted).

Levin, Stewart A. and Fritz Foss, 2014, Downward continuation of Mars SHARAD data: SEG Technical Program Expanded Abstract **33** (Submitted).

Li, Y., Y. Zhang, and J. Claerbout, 2012, Hyperbolic estimation of sparse models from erratic data: Geophysics **77**, 1–9.

Li, Y., P. Shen, and C. Perkins, 2012, VTI migration velocity analysis using RTM: SEG Technical Program Expanded Abstract **31**.

Li, Y., Image-guided WEMVA for azimuthal anisotropy: SEG Technical Program Expanded Abstract **32**.

R.E. Plessix and Y. Li, 2013, Waveform acoustic impedance inversion with spectral shaping: Geophysical Journal International **195**(1), 301–314, 2013.

Li, Y., B. Biondi, D. Nichols, and R. Clapp, 2014, Wave equation migration velocity analysis for VTI models: Accepted for publication in Geophysics, 2014.

Li, Y., M. Wong, and R. Clapp, 2014, Equivalent accuracy by fractional cost: overcoming temporal dispersion: Submitted to SEG 2014.

Li, Y., D. Nichols, and G. Mavko, 2014, Stochastic rock physics modeling for seismic anisotropy: Submitted to SEG 2014.

Li, Y., B. Biondi, R. Clapp, and D. Nichols, 2014, Rock physics constrained anisotropic wave-equation migration velocity analysis: Submitted to SEG 2014.

Maharramov M., and Biondi, B., 2014, Joint full-waveform inversion of time-lapse seismic data sets. SEG Technical Program Expanded Abstracts, 954–959.

Maharramov M., and Biondi, B., 2014, Joint 4DFWI with model-difference regularization. SEG-AGU Summer Research Workshop. Advances in Active+Passive "Full Wavefield" Seismic Imaging: From Reservoirs to Plate Tectonics.

Maharramov M., and Biondi, B., 2015, Robust Simultaneous Time-lapse Full-waveform Inversion with Total-variation Regularization of Model Difference. EAGE Technical Program Expanded Abstracts (accepted).

Maharramov M., 2015, Efficient Finite-difference Modelling of Acoustic Wave Propagation in Anisotropic Media with Pseudo-sources. EAGE Technical Program Expanded Abstracts (accepted).

Maharramov M., and Biondi, B., 2015, Simultaneous TV-regularized time-lapse FWI with application to field data. SEG Technical Program Expanded Abstracts (submitted).

Martin E.R., J. Ajo-Franklin, S. Dou, N. Lindsey, T.M. Daley, B. Freifeld, M. Robertson, A. Wagner, and C. Ulrich, 2015, Interferometry of ambient noise from a trenched distributed acoustic sensing array. SEG Technical Program Expanded Abstracts.

Ajo-Franklin, J., N. Lindsey, S. Dou, T.M. Daley, B. Freifeld, E.R. Martin, M. Robertson, C. Ulrich, and A. Wagner, 2015, A field test of distributed acoustic sensing for ambient noise recording. SEG Technical Program Expanded Abstracts.

Li, Y., H. Yang, E.R. Martin, K. Ho, and L. Ying, 2015, Butterfly factorization. Multiscale Modeling and Simulation, 13, 714–732.

Shen, Y., B. Biondi, R. Clapp, and D. Nichols, 2013, Wave-equation migration Q analysis (WEMQA): EAGE Workshop on Seismic Attenuation Extended Abstract 2013.

Shen, Y., B. Biondi, R. Clapp, and D. Nichols, 2014, Wave-equation migration Q analysis (WEMQA): SEG Technical Program Expanded Abstracts.

Li, Y., Y. Shen and P. Kang, 2015, Integration of seismic and fluid-flow data: a two-way road linked by rock physics :77th EAGE Conference and Exhibition.

Wong, M., and Ronen, S., and Biondi, B.L., 2012, Joint imaging with streamer and ocean bottom data. SEG Technical Program Expanded Abstracts, 1–5.

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

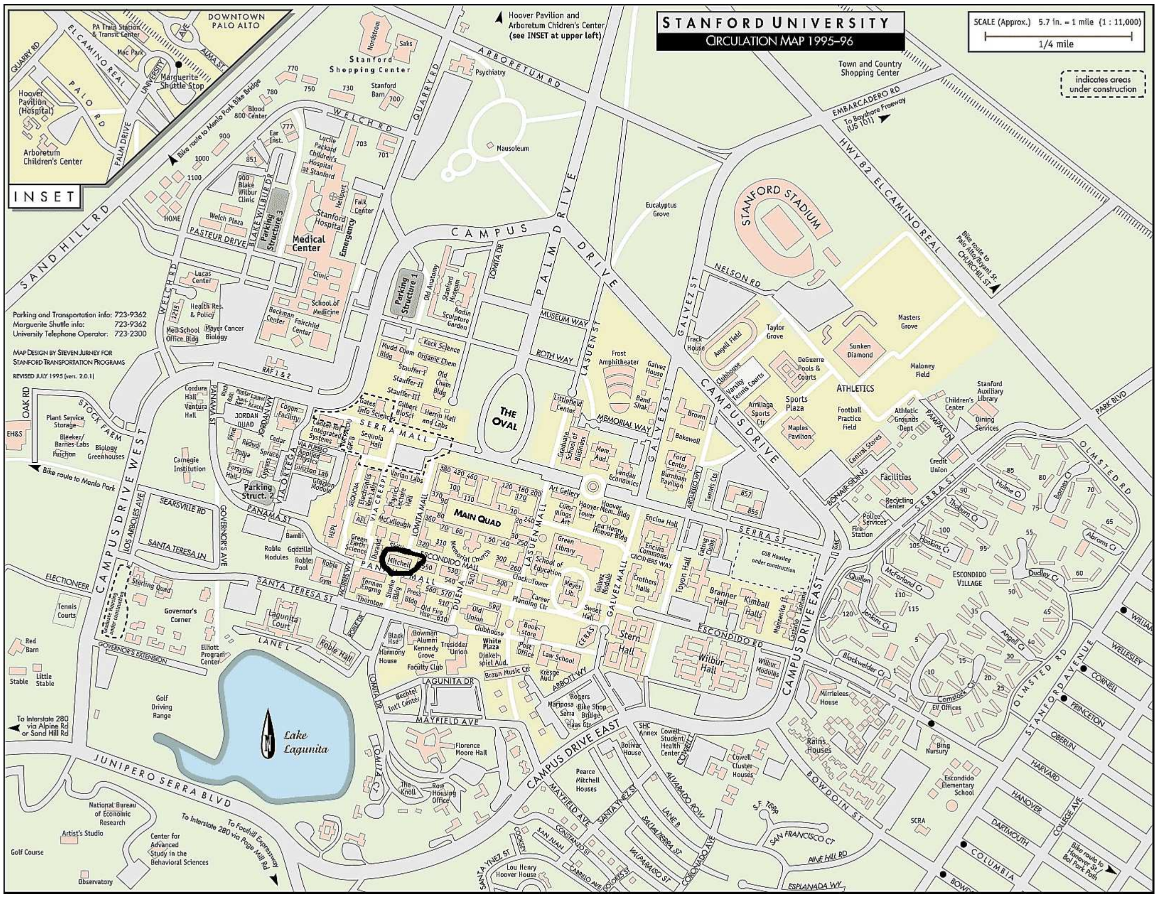
Wong, M., 2013, Handling salt reflection in Least-squares RTM SEG Technical Program Expanded Abstracts, 3921–3925.

Wong, M., B. Biondi, S. Ronen, C. Perkins, M. Merritt, V. Goh, and R. Cook, 2014, Robbust least squares RTM on the 3D Deimos ocean bottom node dataset SEG Technical Program Expanded Abstracts, (Submitted)

Wong, M., B. Biondi, and S. Ronen, 2014, Imaging with multiples using least-squares reverse-time migration. The Leading Edge, (submitted)

Huy Le and Stewart A. Levin, 2014, Removing shear artifacts in acoustic wave propagation in orthorhombic media: SEG Technical Program Expanded Abstracts, 486–490.

Huy Le, Biondo Biondi, Robert G. Clapp, and Stewart A. Levin, 2015, Using a nonlinear acoustic wave equation for anisotropic inversion: SEG Technical Program Expanded Abstracts (Submitted).


STANFORD UNIVERSITY
CIRCULATION MAP 1995–96
SCALE (Approx.) 5.7 in. = 1 mile (1 : 11,000)
1/4 mile
indicates areas under construction
Hoover Pavilion and Arboretum Children's Center (see INSET at upper left)
INSET
DOWNTOWN PALO ALTO
PA Train Station & Transit Center
Marguerite Shuttle Stop
Hoover Pavilion (Hospital)
Arboretum Children's Center
Parking and Transportation info: 723-9362
Marguerite Shuttle info: 723-9362
University Telephone Operator: 723-2300
MAP DESIGN BY STEVEN JURNEY FOR STANFORD TRANSPORTATION PROGRAMS
REVISED JULY 1995 (vers. 2.0.1)
SAND HILL RD
CAMPUS DRIVE
CAMPUS DRIVE WEST
CAMPUS DRIVE EAST
PALM DRIVE
EL CAMINO REAL
JUNIPERO SERRA BLVD
Stanford Shopping Center
Medical Center
Stanford Hospital
Lucile Packard Children's Hospital at Stanford
Mausoleum
Eucalyptus Grove
STANFORD STADIUM
Masters Grove
Sunken Diamond
ATHLETICS
Maples Pavilion
THE OVAL
MAIN QUAD
Memorial Church
Green Library
Hoover Tower
Mitchell
Tresidder Union
Stern Hall
Wilbur Hall
ESCONDIDO VILLAGE
Lake Lagunita
Golf Driving Range
Roble Hall
Lagunita Court
Florence Moore Hall
Governor's Corner
Bike route to Menlo Park
To Interstate 280 via Alpine Rd or Sand Hill Rd
To Interstate 280 via Page Mill Rd
To Foothill Expressway
To Bayshore Freeway (US 101)
Bike route to Palo Alto/Bryant St.
Bike route to Homer St./Bol Park Path