Data-domain and image-domain wavefield tomography

ESTEBAN DÍAZ AND PAUL SAVA, Colorado School of Mines
TONGNING YANG, BP America

Waveform inversion (FWI) requires a good starting model and/or data at low frequency (<1 Hz) for convergence. However, this is not a necessary condition, but an artifact of the objective function defined using differences of observed and simulated data. Image-domain tomographic methods using the same wavefields and wave equations can converge to a reasonable solution from poor starting models and without long-offset and/or low-frequency data. Cascading image-domain and data-domain wavefield tomography eliminates the need for extremely low-frequency in the acquired data.

Introduction

Wavefield tomography (WT) techniques (Tarantola, 1984; Woodward, 1992; Pratt, 1999; Sava and Biondi, 2004a,b; Shen et al., 2003; Biondi and Symes, 2004) have the goal of building accurate subsurface velocity models using wavefields rather than rays. Wavefields are preferable to rays because they are more consistent with the band-limited nature of the recorded data. The accuracy obtained with WT is necessary in today's exploration world where the geologic settings are becoming more challenging, especially in deep marine environments where the drilling costs are high.

Wavefield tomography can be formulated in many different ways, which may appear at first sight to be different from each other. In this article, we analyze the main implementations and highlight their significant differences and similarities.

The data-domain formulation (Lailly, 1983; Tarantola, 1984; Pratt, 1999) is regarded as the methodology with highest resolution. It is conventionally assumed that the high resolution derives from the fact that this method exploits most of the information contained in the observed seismic waveforms. In practice, this implies that we match the observed seismic data with simulated data.

This match occurs only if three conditions are fulfilled:

- the source function used to simulate the wavefields is consistent with the source used in the field to acquire data;
- the physics used to simulate wavefields is consistent with the physics of wave propagation in the Earth;
- the model parameters used to describe the Earth are accurate.

Such stringent requirements are difficult to satisfy in practice, mainly because of the complexity of the physical relations governing wave propagation in the Earth. It is generally understood that seismic waves propagate in a (poro)elastic anisotropic material controlled by a large number of elastic parameters (the stiffness tensor) and by density. However, it is often assumed for computational reasons that wave propagation is controlled by simpler physics, thus reducing the wave equation needed to simulate the seismic waves to acoustic, isotropic, and constant density. Sometimes variable density is also included in the model, not as an independent parameter, but linked to the velocity through an empiric relation like Gardner's (Gardner et al., 1974). It is difficult to match the field observations (elastic, anisotropic, variable density) with simulations (acoustic, isotropic, constant density). Therefore, a significant part of WT is dedicated to removing the components of the observed data that are inconsistent with the assumptions made about the wave equation used for wavefield simulation.

A key component of wavefield tomography (WT) is the objective function (OF) measuring the match between simulated and recorded data. It is assumed that the OF is convex, thus enabling its minimization using gradient-based techniques. In this framework, the model is updated iteratively using a relation like

$$m_{k+1} = m_k + \alpha \Delta m_k, \tag{1}$$

where $\Delta \boldsymbol{m}_k$ is a change of the model that minimizes the mismatch between the simulated and observed wavefields. In this article, we parameterize the model with the squared P-wave slowness.

The model change is obtained from the scaled gradient of the OF which evaluates the mismatch between the observed and simulated wavefields. This calculation is often based on the adjoint state method (Tarantola, 1984; Lailly, 1983). This method consists of four steps: (1) compute the state variables, i.e., seismic wavefields obtained from the source by forward modeling; (2) compute the adjoint source based on the OF and the state variables; (3) compute the adjoint state variables (i.e., seismic wavefields obtained from the adjoint source by backward modeling); and (4) compute the gradient using the state and adjoint state variables.

The state variables are supposed to facilitate interaction between the known measurements (source wavelet and observed data) and the model parameters. Thus, WT state variables are simply the wavefields $u_s(e,\mathbf{x},\omega)$ and $u_r(e,\mathbf{x},\omega)$ obtained by solving a wave equation given the source wavelet $f_s(e,\mathbf{x},\omega)$ and the observed data $f_r(e,\mathbf{x},\omega)$:

$$\begin{pmatrix} \mathcal{L}(\mathbf{m}) & 0 \\ 0 & \mathcal{L}^*(\mathbf{m}) \end{pmatrix} \begin{pmatrix} u_s \\ u_r \end{pmatrix} = \begin{pmatrix} f_s \\ f_r \end{pmatrix}. \tag{2}$$

$\mathcal{L}(\mathbf{x},\omega,\mathbf{m})$ and $\mathcal{L}^*(\mathbf{x},\omega,\mathbf{m})$ are the forward and backward wave-equation operators, respectively; $\mathbf{m}$ are the model parameters, e is the experiment index (shot index, plane-wave index, encoding index), ω is the angular frequency, and $\mathbf{x}$ are space coordinates (x,y,z). It is apparent that u_s and u_r are the same wavefields used in conventional migration (e.g., reverse time migration), because they are based on the same wave equation and on the same input data.

The adjoint state variables, $a_s(e,\mathbf{x},\omega)$ and $a_r(e,\mathbf{x},\omega)$, are also wavefields simulated using the same wave equation as the state variables:

$$\begin{pmatrix} \mathcal{L}^*(\mathbf{m}) & 0 \\ 0 & \mathcal{L}(\mathbf{m}) \end{pmatrix} \begin{pmatrix} a_s \\ a_r \end{pmatrix} = \begin{pmatrix} g_s \\ g_r \end{pmatrix}. \quad (3)$$

One difference between the state and adjoint state variables is the direction of time flow. If time flows forward for a state variable, then time flows backward for the corresponding adjoint state variable. This is accomplished mathematically by employing the forward and adjoint wave operators, $\mathcal{L}(\mathbf{m})$ and $\mathcal{L}^*(\mathbf{m})$, respectively. A second difference between the state and adjoint state variables is the source functions. For state variables, we use the source wavelet and the observed data, respectively, but for the adjoint state variables we use so-called adjoint sources which are derived from the OF defined using the state variables.

Different OFs can be used in the adjoint source calculation. These OFs are different based on the domain in which they are implemented (i.e., data domain or image domain) and based on the mathematical operation used to compare the source and receiver wavefields.

In this work, we explore and compare the properties of different objective functions and their potential to solve different components of the velocity model. We show how the combination of image-domain tomography, followed by FWI can lead to models that satisfy both the long- and short-wavelength components of the model. The good convergence properties of image-domain tomography provide a better starting point for high-resolution methods than those posed in the data domain.

Data-domain wavefield tomography

We can formulate data-domain WT (dWT) as an inverse problem based on an OF defined using the difference between the source and receiver wavefields (Tarantola, 1984; Pratt, 1999; Sirgue and Pratt, 2004; Plessix, 2006; Vigh and Starr, 2008; Virieux and Operto, 2009):

$$J_D = \sum_e \left\| K_D(u_s - u_r) \right\|^2_{\mathbf{x},\omega}. \quad (4)$$

Here, $K_D(e,\mathbf{x})$ is a mask operator restricting the wavefields to the receiver positions (i.e., to the observed and simulated data). In practice, this OF is used directly based on the simulated data $K_D u_s(e,\mathbf{x},\omega)$ and observed data $K_D u_r(e,\mathbf{x},\omega)$; only the source wavefield reconstruction is necessary. However, we use the notation based on wavefields instead of data in order to highlight similarities with other techniques later in the article.

This common OF suffers, among other things, from cycle skipping because of the oscillatory nature of the subtracted wavefields. For illustration, consider the synthetic data in Figures 1a and 1b. The red and blue lines are "observed" and "simulated" data, respectively. In this example, we consider low and high frequency data for the same amount of time delay. The data in Figure 1a are different but within a cycle of one-another; however, the data in Figure 1b skip several cycles between corresponding portions of the wavelet. As illustrated in Figure 2a, cycle skipping depends on the frequency of the data, as well as on the relative delay between the data. In turn, the relative delay between the simulated and observed data depends on the accuracy of the model, as well as on the total propagation distance over which such delays can be accumulated. Figure 2a shows the dependence of the OF with respect to frequency and delay, and highlights the existence of local minima which are a direct consequence of data cycle skipping. On the other hand, the difference OF is steep in the vicinity of the global minimum, thus indicating that if we are in the appropriate basin of convergence, we can obtain models with high resolution.

The cycle-skipping problem is usually addressed by bootstrapping the frequency from low to high (Bunks et al., 1995; Sirgue and Pratt, 2004), thus requiring that low-frequency data (~1 Hz) are acquired in the field. If this condition is satisfied, we can define the adjoint source, $g_s(e,\mathbf{x},\omega)$, based on the source and receiver wavefields

$$g_s = K_D \overline{K_D}(u_s - u_r), \quad (5)$$

compute the adjoint state variable, $a_s(e,\mathbf{x},\omega)$, using backward modeling,

$$\mathcal{L}^* a_s = g_s, \quad (6)$$

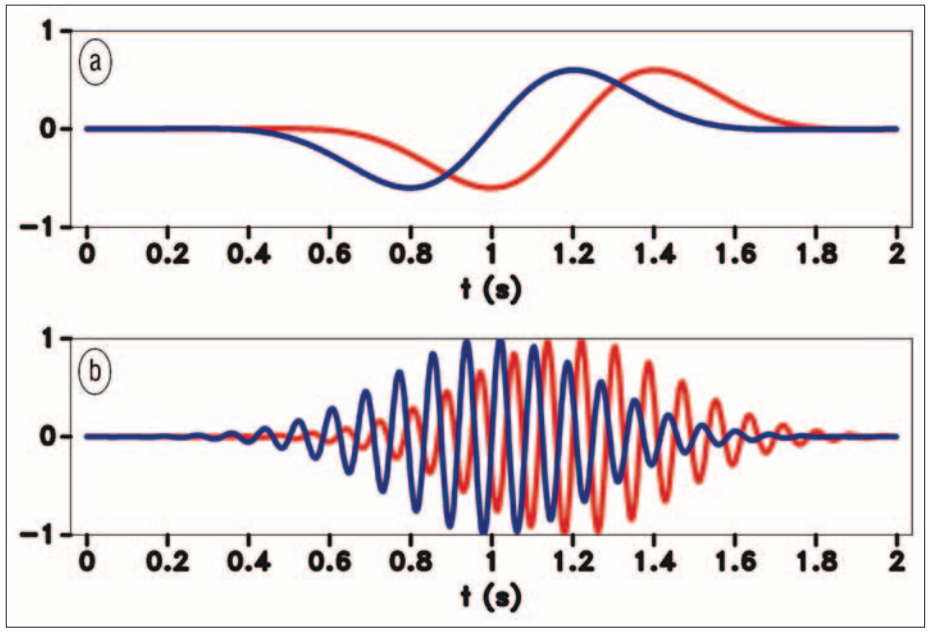

Figure 1. *(a) Low-frequency data and (b) high-frequency data. Blue and red correspond to observed and simulated data.*

and evaluate the OF gradient $\nabla_m J_D(\mathbf{x})$ by correlating the state (us) and adjoint state (as) variables (Plessix, 2006):

$$\nabla_{\mathbf{m}} J_D(\mathbf{x}) = \sum_{e,\omega} \omega^2 \left(\overline{u_s} a_s\right). \quad (7)$$

The data-domain gradient is simply the correlation of the source wavefield and a receiver wavefield constructed based on the difference between the observed and simulated data.

Alternatively, instead of requiring low-frequency data, we can define a different OF based on correlation of the source and receiver wavefields:

$$J_D = \sum_e \left\| K_D P(u_s \star u_r) \right\|^2_{\mathbf{x},\omega,\tau}, \quad (8)$$

where τ is the correlation time lag. If the two wavefields are in phase (i.e., they match), then their correlation is maximum at zero lag, $\tau = 0$. The penalty function $P(\tau)$ is designed to nullify the correlation at zero time lag, and preserve the correlation at nonzero time lag which indicates model inaccuracy. Unlike the difference OF, the correlation OF does not suffer from cycle skipping (Figure 2b) and therefore enables us to converge to the OF global minimum even if the starting time delay between the source and receiver wavefields is large. On the other hand, the correlation OF is fairly flat near the global minimum, which indicates that convergence is slow and does not lead to high-resolution models. Furthermore, even if the model is correct, the correlation OF does not drop to zero because we are operating with band-limited signals of fairly low frequency band. Alternative OFs, e.g., using deconvolution (Luo and Sava, 2011), might partially address this problem, but we do not elaborate further on this distinction.

It appears that the difference and correlation OFs have complementary properties: difference (sensitive to cycle-skipping, high-resolution models); correlation (insensitive to cycle-skipping, low-resolution models). Therefore, it would be tempting to seek OFs that combine the properties of the difference and correlation. One straightforward possibility is to use a linear combination of the difference and correlation OFs:

$$J_M = (1-c)\, J_D + c\, J_C \quad (9)$$

where $0 \le c \le 0$ is a tuning parameter balancing the contribution of the different OFs. Aside from the obvious question about how we should balance the weights assigned to the two components of this mixed OF, its convergence properties when starting with an inaccurate model are far from guaranteed. As seen in Figure 2c, this combined OF is still multimodal and does not ensure convergence to a global minimum regardless of frequency band and of the starting model (i.e., we may still be trapped in a local minimum). Different implementations of this idea (Fleury and Perrone, 2012) may address this convergence problem by allowing the parameter c to change gradually from 1 to 0 (i.e., transition from the correlation to the difference OF during iterations).

Image-domain wavefield tomography

Data-domain WT implementation is not the only possibility. We can consider alternative implementations in which we compare the wavefields through correlation everywhere in the imaged space (i.e., we can use image-domain methods

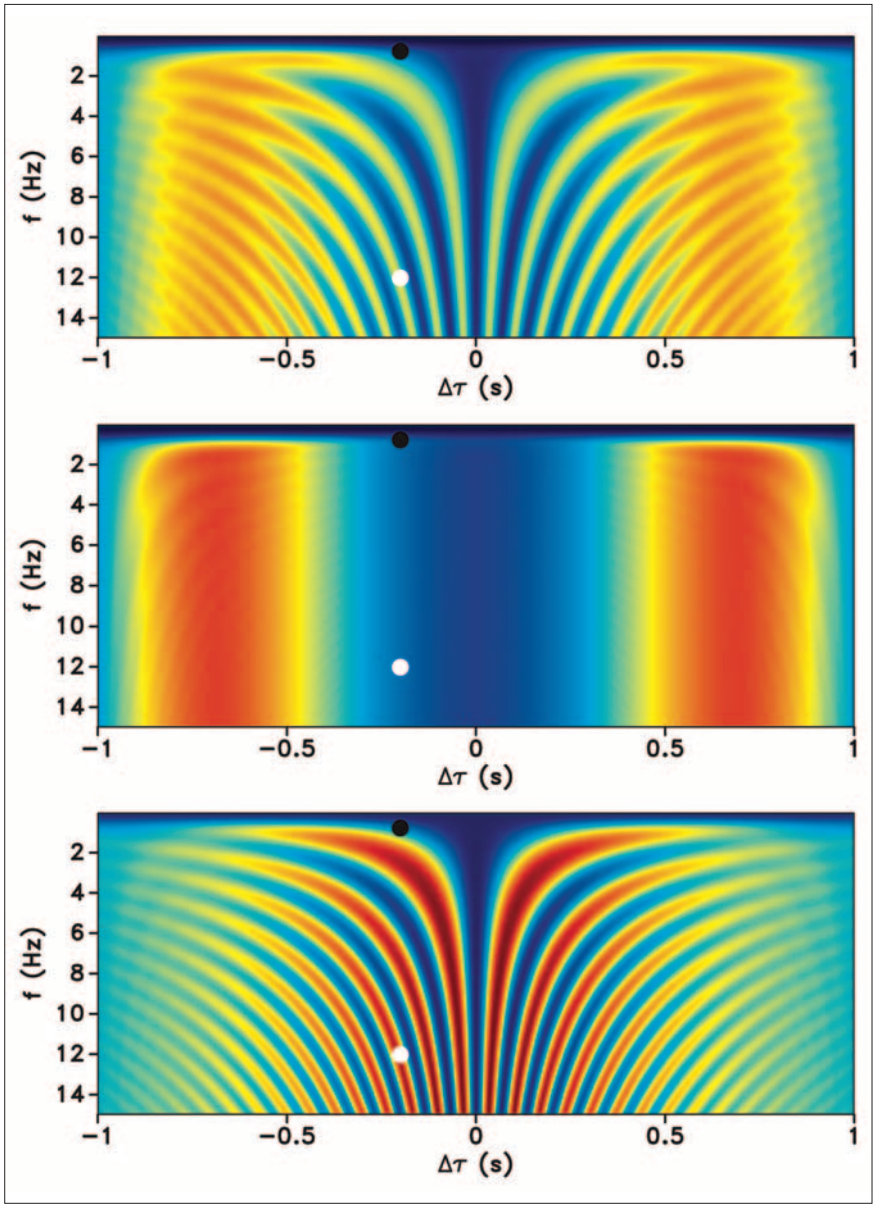

Figure 2. *(a) Difference, (b) correlation, and (c) mixed OFs. The black and blue dots correspond to the low- and high-frequency data, respectively.*

in which we can correlate the wavefields not only as a function of time lags, but also as a function of space lags). In this case, we can construct local crosscorrelations for which we preserve the correlation lags both in time and in space. This construction also allows us to take into account the directionality of waves propagating in the medium under investigation. This is, of course, just conceptual, because local crosscorrelations at every location and all times are computationally intensive. Furthermore, in this case we do not know where the source and receiver wavefields should be similar to each other, as is the case for the data-domain methods when we know that the correlation should be maximum at the receiver locations. In this case, we need to rely on the "semblance principle", which states simply that the model is accurate when the images constructed from different experiments are similar to one-another. In other words, we can make use of so-called extended images, $r(\mathbf{x},\lambda,\tau)$, (Sava and Vasconcelos, 2011)

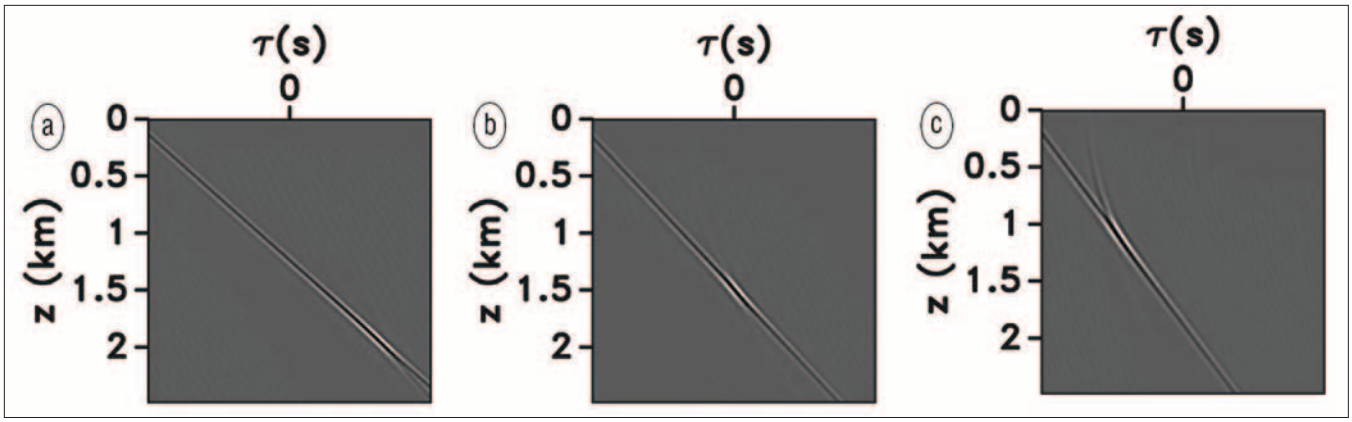

Figure 3. *Time-lag extended images for low, correct, and high velocity, respectively.*

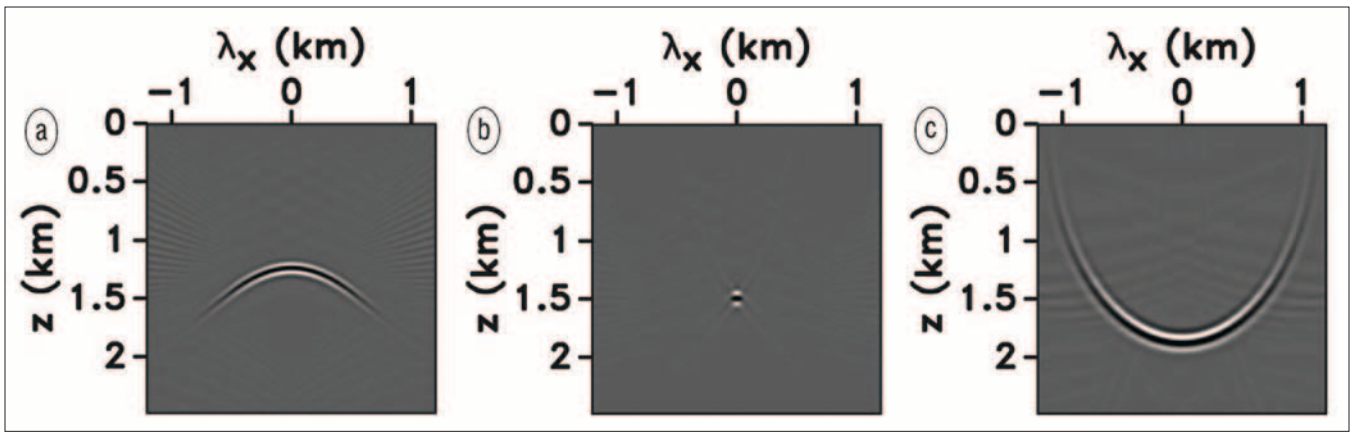

Figure 4. *Space-lag extended images for low, correct, and high velocity, respectively.*

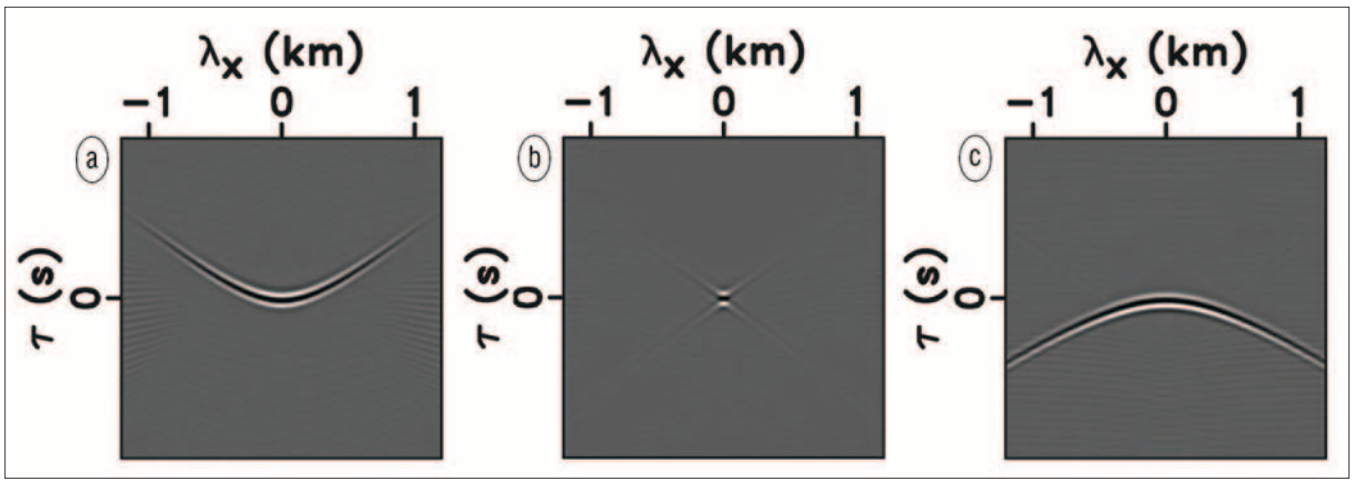

Figure 5. *Space/time-lag extended images for low, correct, and high velocity, respectively.*

$$r = \sum_{\omega} \overline{T(-\lambda)u_s} T(\lambda) u_r e^{2i\omega\tau}. \quad (10)$$

$T(\lambda)$is a space shift operator such that $T(\lambda)u(\mathbf{x},\omega) = u(\mathbf{x}+\lambda, \omega)$ and λ and τ are cross-correlation lags in space and time. As discussed by Sava and Vasconcelos (2011), extended images are sensitive to model inaccuracy, as can be seen in Figures 3a–c for time-lag gathers, in Figures 4a–4c for space-lag gathers and in Figures 5a–c for space/time-lag gathers.

Image-domain WT (iWT) using extended imaging is based on the OF

$$J_I = \left\| K_I P r \right\|^2_{\mathbf{x},\lambda,\tau}, \quad (11)$$

where $r(\mathbf{x},\lambda,\tau)$ are extended images, the mask $K_I(\mathbf{x})$ restricts the evaluation of the OF to some image locations (e.g., to common-image gathers) and $P(\lambda,\tau)$ is a penalty operator applied in the extended space (Shen and Symes, 2008; Symes, 2009; Yang and Sava, 2012). The extended image $r(\mathbf{x},\lambda,\tau)$ measures the similarities around the location $\mathbf{x}$ between the source and receiver wavefields. This OF does not suffer from the cycle-skipping problem, because it is based on the correlation of the source and receiver wavefields, instead of their difference.

The source and receiver adjoint sources, $g_s(e,\mathbf{x},\omega)$ and $g_r(e,\mathbf{x},\omega)$, can be written as (Yang and Sava, 2012)

$$\begin{pmatrix} g_s \\ g_r \end{pmatrix} = \begin{pmatrix} K_I \overline{K_I} \sum_{\tau,\lambda} T(\lambda)(P\bar{P}r)T(\lambda)\overline{u_r} e^{-2i\omega\tau} \\ K_I \overline{K_I} \sum_{\tau,\lambda} T(-\lambda)(P\bar{P}r)T(-\lambda)u_s e^{-2i\omega\tau} \end{pmatrix} \quad (12)$$

and are used to simulate the adjoint state variables, as discussed earlier (Equation 3). The image-domain OF gradient $\nabla_m J_I(\mathbf{x})$ is the correlation of the state (u_s, u_r) and adjoint state (a_s, a_r) variables: (Plessix, 2006):

$$\nabla_{\mathbf{m}} J_I = \sum_{e,\omega} \omega^2 \left(\overline{u_s} a_s + \overline{u_r} a_r \right) \quad (13)$$

Discussion

Both dWT and iWT use the same wavefields and wave operators, thus describing two forms of WT which differ essentially just in the definition of the OF. The image-domain OF is smooth and allows convergence from a poor starting model, even if low frequencies (i.e., ≤ 1 Hz) are not present in the data – we avoid cycle skipping. The data-domain OF is more abrupt near the correct model, which enables convergence to a good quality model—we achieve high resolution. Therefore, dWT and iWT complement each other and can be cascaded to obtain model updates even when the starting model is wrong and when low-frequency data are not recorded. We conclude that the requirement that low-frequency data are necessary for WT is simply an artifact of the definition of the data-domain OF. Using alternative OFs in the image-domain, we can construct models that are close enough to enable convergence of data-domain WT in the more conventional seismic band.

As indicated earlier, the data-domain gradient is simply an image of the difference between observed and simulated data. This gradient is sharp (image-like) for the reflected component of the data, but it is smooth for the transmitted component of the data (diving waves). For reflected data, the data-domain gradient contains a minor smooth component caused by the correlation between wavefields propagating in the same direction (Xu et al., 2012; Dıaz and Sava, 2012). The same is true in the image-domain where the gradient contains a smooth component, as well as a sharp component caused by the correlation between wavefields propagating in opposite directions.

We also note that model regularization with a-priori information can constrain the model updates, as well as reduce the size of the model space

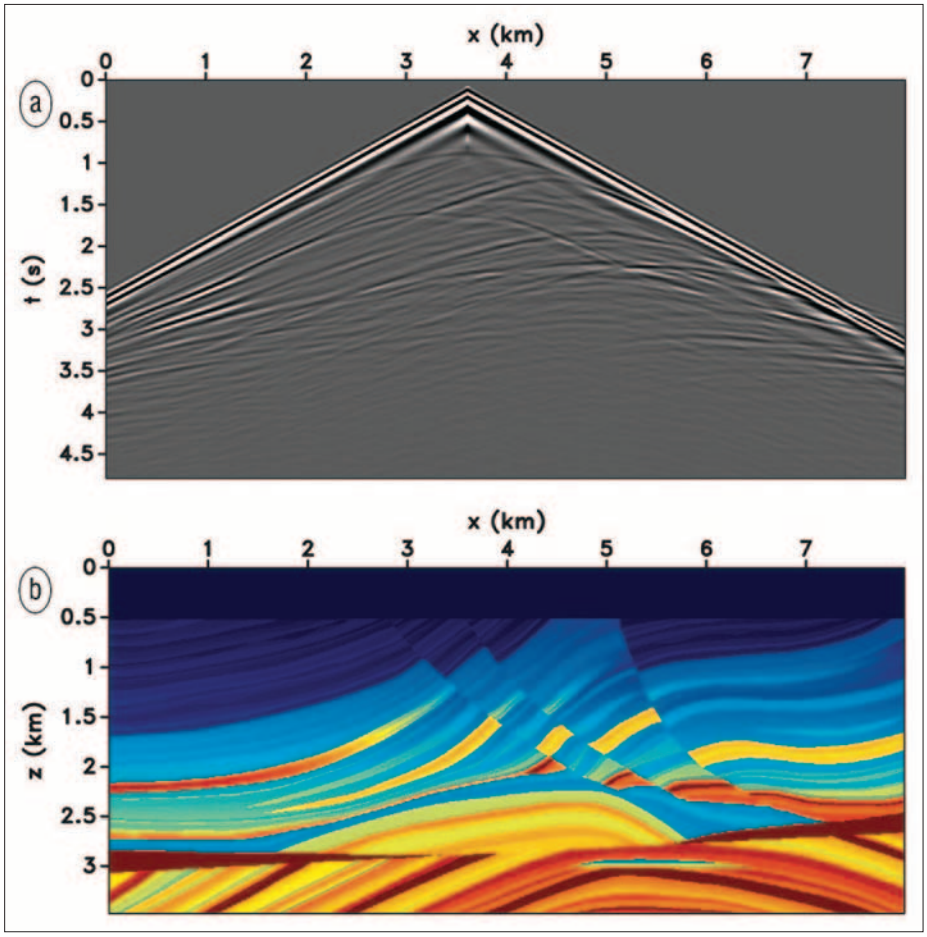

Figure 6. *(a) A shot with short acquisition aperture (no diving waves), and (b) the Marmousi model.*

(Shen and Symes, 2008; Ma et al., 2012; Williamson et al., 2011; Guitton et al., 2012). For example, we could use geologic knowledge to shape the updated models, or we could use well information to constrain the magnitude of the update. This process can be applied to either the data- or the image-domain WT, because both techniques describe similar inverse problems and constrain the same Earth model.

Example

We illustrate the cascaded WT method with data (Figure 6a) simulated in the Marmousi model (Figure 6b), using relatively short offsets and in a frequency band above 3 Hz (Figure 7a), thus eliminating offsets and frequencies that enable a successful start of conventional dWT. Our dWT is initialized either with a model or with a model obtained by iWT (Figure 8a). Figures 8b and 8c show the dWT results using the two starting models, respectively. Both starting models generate reasonable results in the upper part of the model, but only the iWT model is sufficiently close to allow a good reconstruction in the deeper part.

Conclusion

Wavefield tomography can be formulated in the data-domain or in the image-domain based on the same wavefields and wave equation. Image-domain wavefield tomography can construct good starting models for the data-domain approach, because its OF is not sensitive to cycle skipping and its characterized by good conversion properties. Conversely, data-domain wavefield tomography can construct high-resolution models, because this OF is steep in the vicinity of the global minimum. The image-domain wavefield tomography step can produce models that serve as good starting models for the data-domain approach. Therefore, cascading both approaches alleviates the need for extremely low frequencies in the acquired data, and leads to robust methodology that can be used with inaccurate starting models. TLE

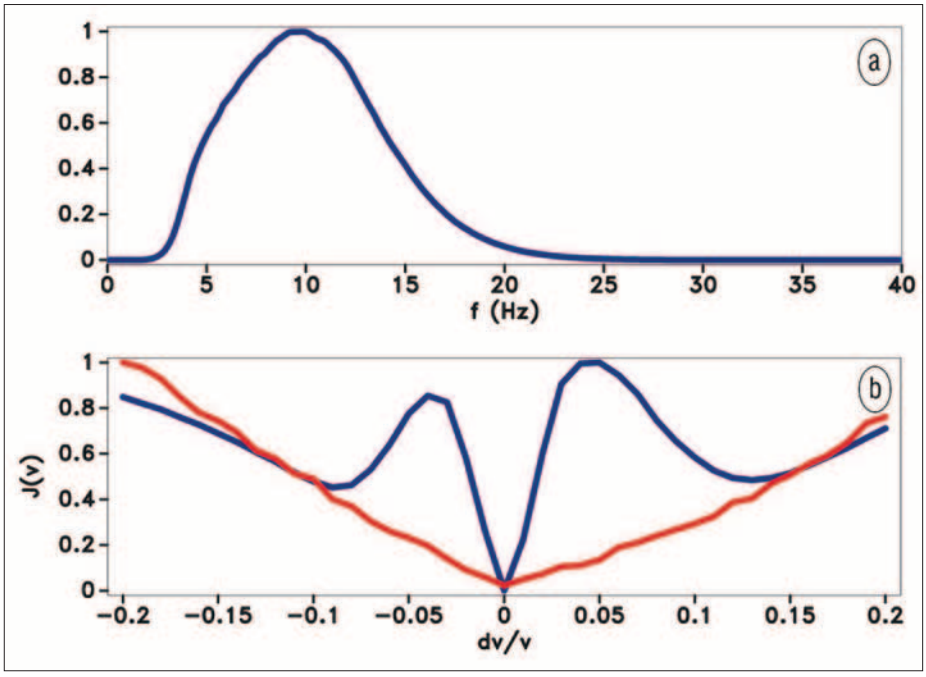

Figure 7. *(a) The data spectrum, and (b) the OF for dWT (blue) and iWT (red).*

References

Biondi, B. and W. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: Geophysics, **69**, 1283–1298, http://dx.doi.org/10.1190/1.1801945.

Bunks, C., F. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473, http://dx.doi.org/10.1190/1.1443880.

Díaz, E. and P. Sava, 2012, Understanding the reverse time migration backscattering: noise or signal?: 82nd Annual International Meeting, SEG, Expanded Abstracts, http:dx.doi.org/10.1190/segam2012-1203.1.

Fleury, C. and F. Perrone, 2012, Bi-objective optimization for the inversion of seismic reflection data: Combined fwi and mva: 82nd Annual International Meeting, SEG, Expanded Abstracts, http:dx.doi.org/10.1190/segam2012-0797.1.

Gardner, G., L. Gardner, and A. Gregory, 1974, Formation velocity and density the diagnostic basics for stratigraphic traps: Geophysics, **39**, 770–780, doi:http:dx.doi.org/ 10.1190/1.1440465.

Guitton, A., G. Ayeni, and E. Díaz, 2012, Constrained full-waveform inversion by model reparameterization: Geophysics, **77**, no. 2, R117–R127, http://dx.doi.org/10.1190/geo2011-0196.1.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations, *in* J. Bednar, ed., Conference on Inverse Scattering: Theory and Application: Society for Industrial and Applied Mathematics, 206–220.

Luo, S. and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion: 81st Annual International Meeting, SEG, Expanded Abstracts, 2788–2792, http://dx.doi.

org/10.1190/1.3627773.

Ma, Y., D. Hale, B. Gong, and Z. Meng, 2012, Image-guided sparse-model full waveform inversion: Geophysics, **77**, no. 4, R189–R198, http://dx.doi.org/10.1190/GEO2011-0395.1.

Plessix, R.-E., 2006, A review of the adjoint state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.

Pratt, R., 1999, Seismic waveform inversion in the frequency domain, part—1: Theory and verification in a physical scale model: Geophysics, **64**, 888–901, http://dx.doi.org/10.1190/1.1444597.

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

Sava, P. and B. Biondi, 2004b, Wave-equation migration velocity analysis—II: Subsalt imaging examples: Geophysical Prospecting, **52**, 607–623.

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

Shen, P. and W. W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, no. 5, VE49–VE59, http://dx.doi.org/10.1190/1.2972021.

Shen, P., W. W. Symes, and C. C. Stolk, 2003, Differential semblance velocity analysis by wave equation migration: 73rd Annual International Meeting, SEG, Expanded Abstracts, 2132–2135, http://dx.doi.org/10.1190/1.1817759.

Sirgue, L. and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248, http://dx.doi.org/10.1190/1.1649391.

Symes, W., 2009, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790.

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

Vigh, D. and E. W. Starr, 2008, 3D prestack plane-wave, full-waveform inversion: Geophysics, **73**, no. 5, VE135–VE144, http://dx.doi.org/10.1190/1.2952623.

Virieux, J. and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26, http://dx.doi.org/10.1190/1.3238367.

Williamson, P., A. Atle, W. Fei, and D. Hale, 2011, Regularization of wave-equation migration velocity analysis by structure-oriented smoothing: 81st Annual International Meeting, SEG, Expanded Abstracts, 3877–3881, http://dx.doi.org/10.1190/1.3628015.

Woodward, M., 1992, Wave-equation tomography: Geophysics, **57**, 15–26, http://dx.doi.org/10.1190/1.3255696.

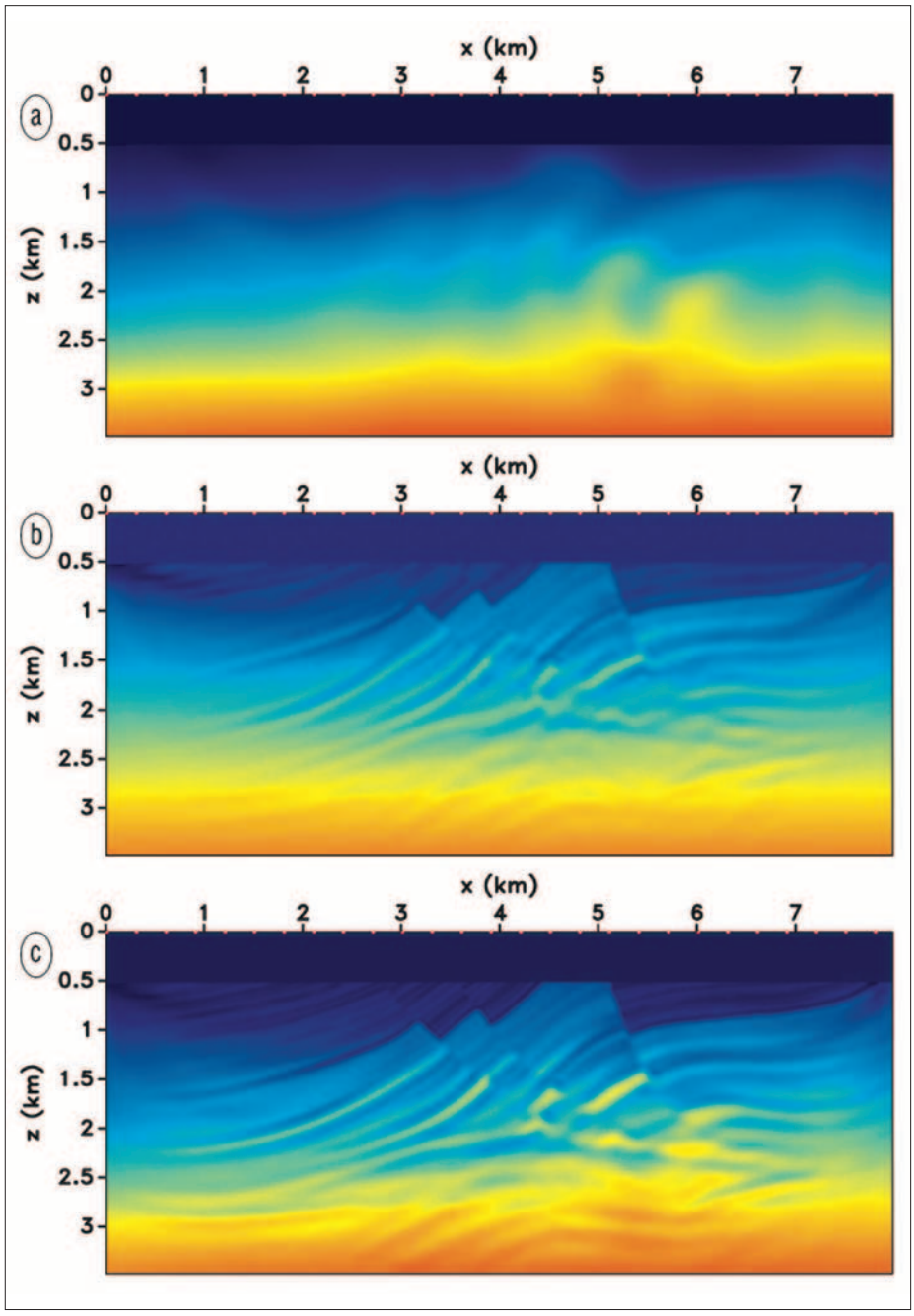

Figure 8. *(a) iWT model starting from a v(z) model, (b) dWT model starting from the same v(z) model, and (c) dWT model starting from the iWT model.*

Xu, S., D. Wang, F. Chen, G. Lambar´e, and Y. Zhang, 2012, Inversion on reflected seismic wave: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: http://dx.doi.org/10.1190/segam2012-1473.1.

Yang, T. and P. Sava, 2012, Image-domain wavefield tomography with extended common-image point gathers: Geophysical Prospecting, submitted for publication.

Acknowledgments: We acknowledge the sponsors of the Center for Wave Phenomena at Colorado School of Mines. The reproducible numeric examples use the Madagascar open-source package (http://www.ahay.org).

Corresponding author: ediazpan@mines.edu