Comparison between time domain and frequency domain least-squares reverse time migration

Lei Yang and Daniel O. Trad

ABSTRACT

In this report, we compare the algorithms of least-squares reverse time migration (LSRTM) in both time and frequency domain and propose a full waveform inversion (FWI) based LSRTM method in the frequency domain. First we show the mathematical equivalence between the gradient of the FWI objective function and the reverse time migration (RTM) imaging condition. Then, we use the FWI formulation with the truncated Newton's method, to solve the linear equation which relates Hessian, model perturbation and the gradient by linear conjugate gradient method. We use simple layer models to compare the two formulations, LSRTM in time and frequency domain. Because of convergence problems that we have not solved yet, we get lower resolution images with the frequency domain FWI-LSRTM method. On the other hand, when the model is inaccurate, the reflector depth seems less affected in the frequency domain. The FWI-based LSRTM method seems to be more robust to velocity errors even if we don't correct the background model as usually done in FWI. Low frequencies seem to be less affected by the inaccurate velocities, and by model smoothness than the high frequencies, suggesting using methods from low frequencies to constraint the high frequencies can help to develop a more robust LSRTM.

INTRODUCTION

The recorded seismic data can be treated as the result of forward modeling problem and this is associated with solving the wave equation. For seismic inversion, the migration operator is adjoint to the forward modeling operator. RTM, as a two-way migration method, migrates the data residual using a zero-lag cross-correlation on the forward and backward propagated wavefields. For conventional LSRTM methods in acoustic medium, the velocity model can be split into two parts: a long-wavelength component, which corresponds to the low frequency feature, and a short wavelength component, which corresponds to the high-frequency feature in the model (Geng and Innanen, 2016). Based on this, the wavefield also consists of two parts: the incident wavefield using wavelets as the source and the scattered wavefield using data residual as the source. By the iterative algorithm of Born modeling and RTM, the reflectivity model is solved by the conjugate gradient method. However, this method depends largely on the initial model, such that if the initial model is inaccurate then the result is also wrong.

Similar to LSRTM, FWI also involves minimizing the misfit function between observed data and synthetic data (Virieux and Operto, 2009). Usually, FWI problems can be solved by a two-loop algorithm: the inner loop is to iteratively solve for the model perturbation and the outer loop is to update the current model and compute the synthetic data to get the new residual. The inner loop can be treated as the LSRTM problem (Chen and Sacchi, 2018). In fact, the gradient of the objective function can be proved to be equal to the image condition of RTM. Therefore, the optimization of the gradient is a way to implement LSRTM. By the Gauss-Newton approximation, using truncated Newton's method is a good way to solve

this problem (Pan et al., 2017). The Hessian-vector product is calculated in each iteration in a matrix-free form. This method seems to be more robust to the inaccuracies in the velocity model although the convergence problems are not solved yet.

THEORY

Frequency domain least-squares reverse time migration

In LSRTM, a reflectivity model is iteratively solved by a forward Born modeling operator and an adjoint (migration) operator. In the frequency domain, the Helmholtz equation is used to solve the forward modeling problem. For 2D acoustic case, the wave equation with constant density can be expressed as:

$$\left(\frac{\omega^2}{\mathbf{v^2(x)}} + \nabla^2\right)\mathbf{u}(\mathbf{x},\omega) = \mathbf{f}(\mathbf{x},\omega), \tag{1}$$

where ω is the angular frequency, ∇^2 is the Laplacian operator, $\mathbf{v}(\mathbf{x})$ is the velocity and $\mathbf{f}(\mathbf{x},\omega)$ represents the source. When the source is the Dirac delta function $\delta(\mathbf{x},\omega)$, the equation 1 changes to:

$$(\frac{\omega^2}{\mathbf{v^2(x)}} + \nabla^2)\mathbf{G}(\mathbf{x},\omega) = \delta(\mathbf{x},\omega), \tag{2}$$

where $\mathbf{G}(\mathbf{x},\omega)$ is the Green's function. To solve equation (1), we use frequency domain finite-difference modeling to discretize the Helmholtz equation. The derivation is based on a 5-point 2D square mesh with constant grid spacing on both x and z direction and we use i and j to represent the horizontal and vertical respectively. By discretizing equation (1), we have

$$\frac{\omega^2}{\mathbf{v^2}}\mathbf{u_{i,j}} + \left(\frac{\mathbf{u_{i-1,j}} - 2\mathbf{u_{i,j}} + \mathbf{u_{i+1,j}}}{\Delta x^2} + \frac{\mathbf{u_{i,j-1}} - 2\mathbf{u_{i,j}} + \mathbf{u_{i,j+1}}}{\Delta z^2}\right) = \mathbf{f_{i,j}} \tag{3}$$

where Δx and Δz are the grid spacing respectively. FIG. 1 shows the element locations for the 5-point 2D finite-difference stencil.

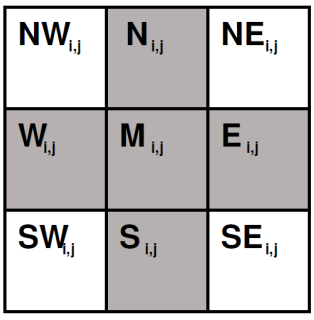


FIG. 1. Symbolic abbreviations for element locations on a 5-point 2D FD stencil (Ajo-Franklin, 2005)

We can write coefficients of the 5-point FD method as:

$$\begin{aligned} M_{i,j} &= \frac{\omega^2}{v_{i,j}^2} - 2\left(\frac{1}{\Delta x^2} + \frac{1}{\Delta z^2}\right) \\ E_{i,j} &= \frac{1}{\Delta x^2} \\ W_{i,j} &= \frac{1}{\Delta x^2} \\ N_{i,j} &= \frac{1}{\Delta z^2} \\ S_{i,j} &= \frac{1}{\Delta z^2} \end{aligned} \tag{4}$$

To avoid boundary reflections, we use absorbing boundary condition (ABC) first proposed by Engquist and Majda (1977)

$$\frac{\partial \mathbf{u}}{\partial \mathbf{n}} - \mathbf{i}\frac{\omega}{\mathbf{v}}\mathbf{u} = \mathbf{0} \tag{5}$$

Take the top boundary$(j = 1)$ for example, the discrete form of ABC is

$$\frac{\mathbf{u_{i,2}} - \mathbf{u_{i,1}}}{\Delta z} + \mathbf{i}\frac{\omega}{\mathbf{v_{i,1}}}\mathbf{u_{i,1}} = 0 \tag{6}$$

FIG. 2 shows the impedance matrix for a 5×5 discrete model and the imaginary entries are just located in the main diagonal of the impedance matrix, which are introduced by ABC.

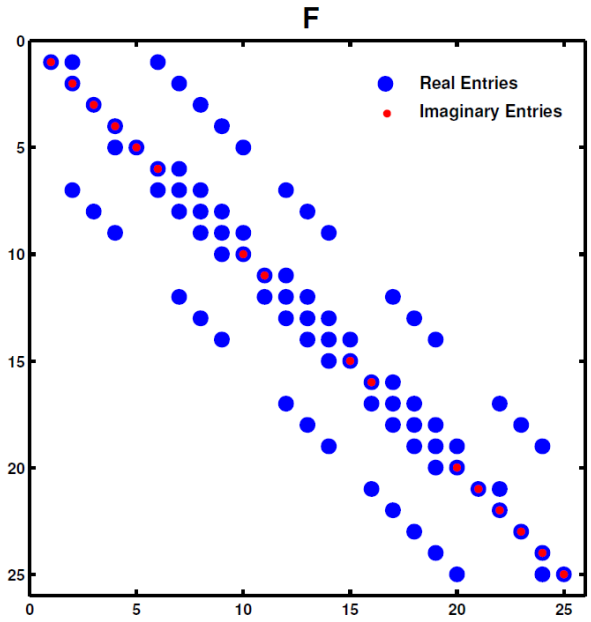


FIG. 2. The impedance matrix for a 5×5 discrete model (Ajo-Franklin, 2005)

In Born modeling, we assume the velocity of the earth model can be split into a smooth part v_0 and a singular part δv:

$$\mathbf{v}(\mathbf{x}) = \mathbf{v_0}(\mathbf{x}) + \delta\mathbf{v}(\mathbf{x}). \tag{7}$$

$\mathbf{v_0}(\mathbf{x})$ represents the long-wavelength components of the velocity model and $\delta\mathbf{v}(\mathbf{x})$ is the short-wavelength velocity model, which contains the singular features. This short-wavelength component will produce reflections and contains high resolution features, which is the main object of the LSRTM. Similarly, the wavefield $\mathbf{u}(\mathbf{x},\omega)$ in the medium $\delta\mathbf{v}(\mathbf{x})$ can be also split into two parts:

$$\mathbf{u}(\mathbf{x},\omega) = \mathbf{u_0}(\mathbf{x},\omega) + \delta\mathbf{u}(\mathbf{x},\omega), \tag{8}$$

where $u_0(x,\omega)$ is the background wavefield and $\delta u(x,\omega)$ is the perturbed wavefield. Using the Taylor's expansion, we have

$$\frac{1}{\mathbf{v^2}(\mathbf{x})} = \frac{1}{\left(\mathbf{v_0}(\mathbf{x}) + \delta\mathbf{v}(\mathbf{x})\right)^2} = \frac{1}{\mathbf{v_0^2}(\mathbf{x})} - \frac{2\delta\mathbf{v}(\mathbf{x})}{\mathbf{v_0^3}(\mathbf{x})} + O(\delta\mathbf{v^2}). \tag{9}$$

If we neglect the higher order term $O(\delta\mathbf{v^2})$, and substitute equation (8) into equation (1),we have

$$\left(\omega^2\left(\frac{1}{\mathbf{v_0^2}(\mathbf{x})} - \frac{2\delta\mathbf{v}(\mathbf{x})}{\mathbf{v_0^3}(\mathbf{x})}\right) + \nabla^2\right)\left(\mathbf{u_0}(\mathbf{x},\omega) + \delta\mathbf{u}(\mathbf{x},\omega)\right) = \mathbf{f}(\mathbf{x},\omega). \tag{10}$$

Rearranging equation (10) and consider $\delta\mathbf{u}(\mathbf{x},\omega)$ is weak, we have the incident wavefield

$$\left(\frac{\omega^2}{\mathbf{v_0^2}(\mathbf{x})} + \nabla^2\right)\mathbf{u_0}(\mathbf{x},\omega) = \mathbf{f}(\mathbf{x},\omega), \tag{11}$$

and scattered wavefield

$$\left(\frac{\omega^2}{\mathbf{v_0^2}(\mathbf{x})} + \nabla^2\right)\delta\mathbf{u}(\mathbf{x},\omega) \approx \omega^2\frac{2\delta\mathbf{v}(\mathbf{x})}{\mathbf{v_0^3}(\mathbf{x})}\mathbf{u_0}(\mathbf{x},\omega), \tag{12}$$

where $\dfrac{2\delta\mathbf{v}(\mathbf{x})}{\mathbf{v_0^3}(\mathbf{x})}$ represents the reflectivity model. The Born approximation is applied in equation (12), where the scattered wavefield $\delta\mathbf{u}(\mathbf{x},\omega)$ is considered very weak and neglected in the right side of equation (12). These two wavefields can be calculated by two finite-difference schemes respectively. For the incident wavefield $\mathbf{u_0}(\mathbf{x},\omega)$, the wavefield is generated by the source $\mathbf{f}(\mathbf{x},\omega)$ in the background model $\mathbf{v_0^2}(\mathbf{x})$ while the source changes to $\omega^2\dfrac{2\delta\mathbf{v}(\mathbf{x})}{\mathbf{v_0^3}(\mathbf{x})}\mathbf{u_0}(\mathbf{x},\omega)$ for the scattered wavefield in the same background model where ω^2 is the second time derivative in the time domain. Applying the ajoint state method on equation (11) and equation (12), we have the RTM imaging condition:

$$\mathbf{m}_{mig}(\mathbf{x}) = \sum_{n_s}\sum_{n_\omega}\frac{\mathbf{1}}{\omega^\mathbf{2}}\mathbf{Re}(\delta\mathbf{u}(\mathbf{x},\omega)\mathbf{G_0^\dagger}(\mathbf{x_s}|\mathbf{x})\mathbf{G_0^\dagger}(\mathbf{x}|\mathbf{x'})\mathbf{f}^\dagger(\mathbf{x},\omega)). \tag{13}$$

$\mathbf{m_{mig}}(\mathbf{x})$ is the migration result $\dfrac{2\delta\mathbf{v}(\mathbf{x})}{\mathbf{v_0^3}(\mathbf{x})}$, which is usually called reflectivity in LSRTM. $\mathbf{G_0^\dagger}(\mathbf{x_s}|\mathbf{x})$ and $\mathbf{G_0^\dagger}(\mathbf{x}|\mathbf{x'})$ represent the conjugate transpose of the Green's functions, which illustrates the cross-correlation of two wavefields as the imaging condition of LSRTM. So far, I have illustrated the forward modeling and migration process in frequency domain

LSRTM. I will use matrix form to show the linearized inversion process in the following section.

The linear modeling step can be expressed as

$$\mathbf{d} = \mathbf{L}\mathbf{m}, \tag{14}$$

and the migration step is

$$\mathbf{m} = \mathbf{L^T}\mathbf{d}, \tag{15}$$

where $\mathbf{d}$ is scattered wavefield data, $\mathbf{L}$ is the modeling operator $\mathbf{G_0}(\mathbf{x_s}|\mathbf{x})\mathbf{G_0}(\mathbf{x}|\mathbf{x}')f(x,\omega)$ and $\mathbf{L}^T$ is the conjugate transpose of $\mathbf{L}$. By using the linear conjugate gradient method, the reflectivity model $\mathbf{m}$ can be iteratively solved by the scheme (Dai et al., 2012):

$$\mathbf{g^{(k)}} = \mathbf{L^T}[\mathbf{L}(\mathbf{m^{(k)}}) - \mathbf{d}], \tag{16}$$

$$\alpha = \frac{(\mathbf{g^{(k)}})^\mathbf{T}\mathbf{g(k)}}{(\mathbf{Lg^{(k)}})^\mathbf{T}\mathbf{Lg^{(k)}}}, \tag{17}$$

$$\mathbf{m^{(k+1)}} = \mathbf{m^{(k)}} - \alpha\mathbf{g^{(k)}}, \tag{18}$$

where α is the step length. This step length is not accurate enough when the forward modeling and migration operator are not exactly adjoint. To improve the convergence rate, a line search method in quasi-linear approach can be applied in this circumstance.

FWI-based LSRTM in frequency domain

In this section, we propose a FWI-based LSRTM algorithm in the 2D acoustic constant density case. This problem is formulated from the objective function of FWI:

$$\mathbf{J}(\mathbf{m}) = \frac{1}{2}\sum_{n_\omega}\sum_{n_s}\|\mathbf{d_{obs}}(\mathbf{x_s},\omega) - \mathbf{d_{syn}}(\mathbf{m},\mathbf{x_s},\omega)\|^2 = \frac{1}{2}\sum_{n_\omega}\sum_{n_s}\|\delta\mathbf{d}\|^2, \tag{19}$$

where $\mathbf{d_{obs}}$ is the observed data, $\mathbf{d_{syn}} = \mathbf{Ru}$ is the synthetic data in which $\mathbf{R}$ is the extraction operator.

The gradient

Starting from equation (19), we will prove that the gradient of FWI objective function is equal to the RTM operator. The squared data residual in equation (19) can be expanded as

$$\begin{aligned}\|\delta\mathbf{d}\|^2 &= (\mathbf{d_{obs}} - \mathbf{d_{syn}})^T(\mathbf{d_{obs}} - \mathbf{d_{syn}})^* \\ &= \mathbf{d_{obs}^T}\mathbf{d_{obs}^*} - \mathbf{d_{obs}^T}\mathbf{d_{syn}^*} - \mathbf{d_{syn}^T}\mathbf{d_{obs}^*} + \mathbf{d_{syn}^T}\mathbf{d_{syn}^*},\end{aligned} \tag{20}$$

where $*$ means conjugate and $\mathbf{T}$ means transpose. Inserting equation (35) into equation (19) and taking derivative of equation (19) with respect to $\mathbf{m}$, we have

$$\begin{aligned}\mathbf{g} &= \frac{\partial \mathbf{J}(\mathbf{m})}{\partial \mathbf{m}} \\ &= -\frac{1}{2}(-\mathbf{d}_{\mathbf{obs}}^{\mathbf{T}}\frac{\partial \mathbf{d}_{\mathbf{syn}}^{*}}{\partial \mathbf{m}} - \mathbf{d}_{\mathbf{obs}}^{*}\frac{\partial \mathbf{d}_{\mathbf{syn}}^{\mathbf{T}}}{\partial \mathbf{m}} + \mathbf{d}_{\mathbf{syn}}^{*}\frac{\partial \mathbf{d}_{\mathbf{syn}}^{\mathbf{T}}}{\partial \mathbf{m}} + \mathbf{d}_{\mathbf{syn}}^{\mathbf{T}}\frac{\partial \mathbf{d}_{\mathbf{syn}}^{*}}{\partial \mathbf{m}}) \\ &= -\frac{1}{2}\left((\mathbf{d_{syn}}-\mathbf{d_{obs}})^{\mathbf{T}}\frac{\partial \mathbf{d}_{\mathbf{syn}}^{*}}{\partial \mathbf{m}} + (\mathbf{d_{syn}}-\mathbf{d_{obs}})^{*}\frac{\partial \mathbf{d_{syn}}}{\partial \mathbf{m}}\right),\end{aligned} \tag{21}$$

Because

$$Z + Z^{*} = \mathbf{Re}Z + i\mathbf{Im}\mathbf{Z} + \mathbf{Re}Z - i\mathbf{Im}Z = 2\mathbf{Re}Z, \tag{22}$$

we can rewrite the gradient as:

$$\begin{aligned}\mathbf{g} &= -\sum_{\mathbf{n}_\omega}\mathbf{Re}\left(\frac{\partial \mathbf{d}_{\mathbf{syn}}^{\mathbf{T}}}{\partial \mathbf{m}}\delta\mathbf{d}^{*}\right) \\ &= -\sum_{\mathbf{n}_\omega}\mathbf{Re}\left(\frac{\partial \mathbf{d}_{\mathbf{syn}}^{\dagger}}{\partial \mathbf{m}}\delta\mathbf{d}\right)\end{aligned} \tag{23}$$

The term $\frac{\partial \mathbf{d}_{\mathbf{syn}}^{\mathbf{T}}}{\partial \mathbf{m}}$ represents the Fréchet derivative and $\delta\mathbf{d}^{*}$ is the conjugate of the data residual. To derive the complete form of the Fréchet derivative, we write the acoustic wave equation in the matrix form:

$$\mathbf{A}(\mathbf{m},\omega)\mathbf{u}(\mathbf{m},\mathbf{x_s},\omega) = \mathbf{f}(\mathbf{x_s},\omega) \tag{24}$$

$$\mathbf{A}(\mathbf{m},\omega) = \left(\omega^{\mathbf{2}}\mathbf{m}(\mathbf{x}) + \nabla^{\mathbf{2}}\right) \tag{25}$$

where $\mathbf{A}(\mathbf{m},\omega)$ is the impedance matrix.

Take derivative on each side of the equation (24) with respect to model parameter $\mathbf{m}$ and put one term in the other side of the equation, we have

$$\mathbf{A}(\mathbf{m},\omega)\frac{\partial \mathbf{u}(\mathbf{m},\mathbf{x_s},\omega)}{\partial \mathbf{m}} = -\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}}\mathbf{u}(\mathbf{m},\mathbf{x_s},\omega). \tag{26}$$

This equation shows that the Fréchet derivative can be obtained by solving the equation (26) with a virtual source term

$$\mathbf{f} = -\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}}\mathbf{u}(\mathbf{m},\mathbf{x_s},\omega) \tag{27}$$

Therefore, the Fréchet derivative can be written as

$$\begin{aligned}\frac{\partial \mathbf{u}(\mathbf{m},\mathbf{x_s},\omega)}{\partial \mathbf{m}} &= -\mathbf{A}^{-\mathbf{1}}(\mathbf{m},\omega)\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}}\mathbf{u}(\mathbf{m},\mathbf{x_s},\omega) \\ &= -\mathbf{G}(\mathbf{x}|\mathbf{x}')\omega^{\mathbf{2}}\mathbf{G}(\mathbf{x_s}|\mathbf{x})\mathbf{f}(\mathbf{x_s},\omega),\end{aligned} \tag{28}$$

Because $\mathbf{d_{syn}} = \mathbf{Ru}$ and $\mathbf{R}$ is a real-valued operator which is independent of model $\mathbf{m}$, the gradient of FWI objective function is

$$\begin{aligned}\mathbf{g} &= -\sum_{\mathbf{n_\omega}}\sum_{\mathbf{n_s}} \mathbf{Re}\left(\frac{\partial \mathbf{d}^\dagger_{\mathbf{syn}}}{\partial \mathbf{m}}\delta\mathbf{d}\right)\\ &= -\sum_{\mathbf{n_\omega}}\sum_{\mathbf{n_s}} \mathbf{Re}\left(\mathbf{R^T}\frac{\partial \mathbf{u}^\dagger(\mathbf{m},\mathbf{x_s},\omega)}{\partial \mathbf{m}}\delta\mathbf{d}\right)\\ &= \sum_{\mathbf{n_\omega}}\sum_{\mathbf{n_s}} \omega^{\mathbf{2}}\mathbf{Re}\left(\mathbf{R^T}\mathbf{G}^\dagger(\mathbf{x}|\mathbf{x}')\mathbf{G}^\dagger(\mathbf{x_s}|\mathbf{x})\mathbf{f}^\dagger(\mathbf{x_s},\omega)\mathbf{R}\delta\mathbf{u}\right)\\ &= \mathbf{R^T R}\sum_{\mathbf{n_\omega}}\sum_{\mathbf{n_s}} \omega^{\mathbf{2}}\mathbf{Re}\left(\mathbf{G}^\dagger(\mathbf{x}|\mathbf{x}')\mathbf{G}^\dagger(\mathbf{x_s}|\mathbf{x})\mathbf{f}^\dagger(\mathbf{x_s},\omega)\delta\mathbf{u}\right).\end{aligned} \tag{29}$$

Comparing the imaging equation (13) and the gradient equation (29), we conclude that the imaging condition of RTM is equal to the gradient of the FWI objective function, except for the small coefficient changes where the coefficient $\sum_{\mathbf{n_\omega}} \frac{\mathbf{1}}{\omega^{\mathbf{2}}}$ changes to $\mathbf{R^T R}\sum_{\mathbf{n_\omega}} \omega^{\mathbf{2}}$. In fact, the term $\mathbf{R^T R}$ is the number of the receivers on the surface.

If we take derivative of the transpose of wavefield $\mathbf{u}(\mathbf{m},\mathbf{x_s},\omega)$ with respect to $\mathbf{m}$ and substitute the result into equation (23), the matrix form of equation (29) can be formed as:

$$\begin{aligned}\mathbf{g} &= -\mathbf{Re}\left(\sum_{\mathbf{n_\omega}}\sum_{\mathbf{n_s}} \mathbf{R^T}\left(-\mathbf{A^{-1}}(\mathbf{m},\omega)\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}}\mathbf{u}(\mathbf{m},\mathbf{x_s},\omega)\right)^{\mathbf{T}}\delta\mathbf{d}^*\right)\\ &= \mathbf{Re}\left(\sum_{\mathbf{n_\omega}}\sum_{\mathbf{n_s}}\left(\mathbf{u}(\mathbf{m},\mathbf{x_s},\omega)^{\mathbf{T}}\left(\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}}\right)^{\mathbf{T}}(\mathbf{A^{-1}}(\mathbf{m},\omega))^{\mathbf{T}}\mathbf{R^T}\delta\mathbf{d}^*\right)\right).\end{aligned} \tag{30}$$

The conjugate of the data residual in the frequency domain is equal to the data time reversed in time domain and the data residual is back projected to the whole space using the operator $\mathbf{R^T}$, before it is propagated back to the subsurface by the term $(\mathbf{A^{-1}}(\mathbf{m},\omega))^{\mathbf{T}}$ (Geng et al., 2017). Since only the real part is considered in the gradient computation, by using the adjoint impedance matrix, we have

$$\mathbf{g} = \mathbf{Re}\left(\sum_{\mathbf{n_\omega}}\sum_{\mathbf{n_s}}\left(\mathbf{u}(\mathbf{m},\mathbf{x_s},\omega)^{\dagger}\left(\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}}\right)^{\dagger}(\mathbf{A^{-1}}(\mathbf{m},\omega))^{\dagger}\mathbf{R}^{\dagger}\delta\mathbf{d}\right)\right), \tag{31}$$

where we keep $\left(\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}}\right)^{\dagger}$ for different parameterization. Therefore, we have the matrix form of gradient computation. Now with the gradient equation, we are able to use steepest decent method to update the initial model in a FWI problem. However, the gradient-based method only accounts for the first-order scattered data, which will cause severe faults when the data contains second-order scattered data. In this situation, the second-ordered derivative of the objective function (Hessian) can alleviate this problem. This requires the Newton-based method to be applied in the model perturbation computation.

The Hessian

We expand the objective function of FWI in Taylor series

$$\mathbf{J}(\mathbf{m}+\delta\mathbf{m}) = \mathbf{J}(\mathbf{m}) + \delta\mathbf{m}^{\mathbf{T}}\mathbf{g} + \frac{\mathbf{1}}{\mathbf{2}}\delta\mathbf{m}^{\mathbf{T}}\mathbf{H}\delta\mathbf{m} + ... \tag{32}$$

When the minimum of the misfit function is reached, and neglecting the higher order terms,

$$\mathbf{J}(\mathbf{m}+\delta\mathbf{m}) - \mathbf{J}(\mathbf{m}) \approx \delta\mathbf{m}^{\mathbf{T}}\mathbf{g} + \frac{\mathbf{1}}{\mathbf{2}}\delta\mathbf{m}^{\mathbf{T}}\mathbf{H}\delta\mathbf{m} = \mathbf{0}. \tag{33}$$

Therefore, we have the linear equation which connects the gradient, Hessian and model perturbation. Starting from this equation, we formulate the linear relationship:

$$\mathbf{H}\delta\mathbf{m} = -\mathbf{g}. \tag{34}$$

Based on this, the objective function of FWI-based LSRTM in image domain is

$$\mathbf{\Phi}(\delta\mathbf{m}) = \frac{\mathbf{1}}{\mathbf{2}}\| - \mathbf{g} - \mathbf{H}\delta\mathbf{m}\|^{\mathbf{2}}. \tag{35}$$

Because Hessian is an extremely large and dense matrix for large scale inverse problem and has large computation cost of its inverse, there are many ways to compute approximate Hessian. If the model dimension is $M \times N$, Hessian will be a $MN \times MN$ symmetric square matrix (MN means $M \times N$). In the following section, I will derive the full Hessian by taking the second derivative of the FWI's objective function. Recall the objective function of FWI

$$\begin{aligned}\mathbf{J}(\mathbf{m}) &= \frac{1}{2}\sum_{n_\omega}\sum_{n_s}\|\mathbf{d_{obs}}(\mathbf{x_s},\omega) - \mathbf{d_{syn}}(\mathbf{m},\mathbf{x_s},\omega)\|^2 \\ &= \frac{1}{2}\sum_{n_\omega}\sum_{n_s}(\mathbf{d_{obs}} - \mathbf{Ru})(\mathbf{d_{obs}} - \mathbf{Ru})^\dagger,\end{aligned} \tag{36}$$

Take first order partial derivative with m for objective function is

$$\frac{\partial\mathbf{J}(\mathbf{m})}{\partial\mathbf{m}} = \frac{\mathbf{1}}{\mathbf{2}}\sum_{\mathbf{n_\omega}}\sum_{\mathbf{n_s}}\left(-\mathbf{R}\frac{\partial\mathbf{u}}{\partial\mathbf{m}}(\mathbf{d_{obs}} - \mathbf{Ru})^\dagger + (\mathbf{d_{obs}} - \mathbf{Ru})(-\mathbf{R}^\dagger\frac{\partial\mathbf{u}^\dagger}{\partial\mathbf{m}})\right) \tag{37}$$

and the second order derivative is

$$\begin{aligned}\frac{\partial\mathbf{J^2}(\mathbf{m})}{\partial\mathbf{m^2}} = \frac{1}{2}\sum_{n_\omega}\sum_{n_s} &-\mathbf{R}\frac{\partial^\mathbf{2}\mathbf{u}}{\partial\mathbf{m^2}}(\mathbf{d_{obs}} - \mathbf{Ru})^\dagger + \frac{\partial\mathbf{u}}{\partial\mathbf{m}}\mathbf{R}\mathbf{R}^\dagger\frac{\partial\mathbf{u}^\dagger}{\partial\mathbf{m}} \\ &+ \frac{\partial\mathbf{u}}{\partial\mathbf{m}}\mathbf{R}\mathbf{R}^\dagger\frac{\partial\mathbf{u}^\dagger}{\partial\mathbf{m}} - \mathbf{R}^\dagger\frac{\partial^\mathbf{2}\mathbf{u}^\dagger}{\partial\mathbf{m^2}}(\mathbf{d_{obs}} - \mathbf{Ru})\end{aligned} \tag{38}$$

Take the real part of equation (38), the full Hessian is

$$\mathbf{H}(\mathbf{m}) = \mathbf{Re}\left(\frac{\partial\mathbf{J^2}(\mathbf{m})}{\partial\mathbf{m^2}}\right) = \mathbf{B}(\mathbf{m}) + \mathbf{C}(\mathbf{m}) \tag{39}$$

where

$$\mathbf{B(m)} = \mathbf{Re}\left(\sum_{\mathbf{n_\omega}}\sum_{\mathbf{n_s}} \frac{\partial \mathbf{u}}{\partial \mathbf{m}}\mathbf{R}\mathbf{R}^\dagger \frac{\partial \mathbf{u}^\dagger}{\partial \mathbf{m}}\right), \tag{40}$$

$$\mathbf{C(m)} = \mathbf{Re}\left(\sum_{\mathbf{n_\omega}}\sum_{\mathbf{n_s}} \mathbf{R}^\dagger(\mathbf{Ru} - \mathbf{d_{obs}})\frac{\partial^\mathbf{2}\mathbf{u}}{\partial \mathbf{m^2}}\right) \tag{41}$$

The matrix $\mathbf{B(m)}$ is known as the Gauss-Newton approximation of the Hessian operator when the expression $\mathbf{C(m)}$ is neglected (Métivier et al., 2013). Substituting equation (28) into (40), the Hessian vector product using Gauss-Newton approximation is

$$\mathbf{H}\delta\mathbf{m} = \mathbf{u}\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}}\mathbf{A^{-1}}(\mathbf{m},\omega)\mathbf{R}\mathbf{R}^\dagger(\mathbf{A^{-1}}(\mathbf{m},\omega))^\dagger\left(\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}}\right)^\dagger \mathbf{u}^\dagger\delta\mathbf{m} \tag{42}$$

To solve the linear equation $\mathbf{H}\delta\mathbf{m} = -\mathbf{g}$, we use linear conjugate gradient method to get $\delta\mathbf{m}$ iteratively.

Time domain least-squares reverse time migration

The 2D acoustic wave equation with constant density in time domain is

$$\frac{\mathbf{1}}{\mathbf{v^2}(\mathbf{x})}\frac{\partial^\mathbf{2}\mathbf{p}(\mathbf{x},\mathbf{t};\mathbf{x_s})}{\partial \mathbf{t^2}} - \nabla^\mathbf{2}\mathbf{p}(\mathbf{x},\mathbf{t};\mathbf{x_s}) = \mathbf{f_s}(\mathbf{x},\mathbf{t};\mathbf{x_s}), \tag{43}$$

where $\mathbf{f_s}(\mathbf{x},\mathbf{t};\mathbf{x_s}) = \mathbf{f}(\mathbf{t}')\delta(\mathbf{x}-\mathbf{x_s})\delta(\mathbf{t}-\mathbf{t}')$ represents the source. Similarly, the wavefield can be split into 2 parts:

$$\left(\frac{1}{\mathbf{v_0^2}(\mathbf{x})}\frac{\partial^\mathbf{2}}{\partial \mathbf{t^2}} + \nabla^2\right)\mathbf{u_0}(\mathbf{x},\mathbf{t}) = \mathbf{f}(\mathbf{x},\mathbf{t}), \tag{44}$$

$$\left(\frac{1}{\mathbf{v_0^2}(\mathbf{x})}\frac{\partial^\mathbf{2}}{\partial \mathbf{t^2}} + \nabla^2\right)\delta\mathbf{u}(\mathbf{x},\mathbf{t}) = \frac{\mathbf{2}\delta\mathbf{v}(\mathbf{x})}{\mathbf{v_0^3}(\mathbf{x})}\frac{\partial^\mathbf{2}\mathbf{u_0}(\mathbf{x},\mathbf{t})}{\partial \mathbf{t^2}}, \tag{45}$$

Equation (44) and (45) represent the incident wavefield and scattered wavefield respectively. Different from frequency domain, we cannot write the wave equation as the matrix multiplication form. Therefore, the wavefield is calculated by each time slice, which can be describe as:

$$\mathbf{p^{n+1}} = \mathbf{2p^n} - \mathbf{p^{n-1}} + \Delta t^2 v^2 \nabla^2 \mathbf{p^n} + \Delta t^2 \mathbf{f^n} \tag{46}$$

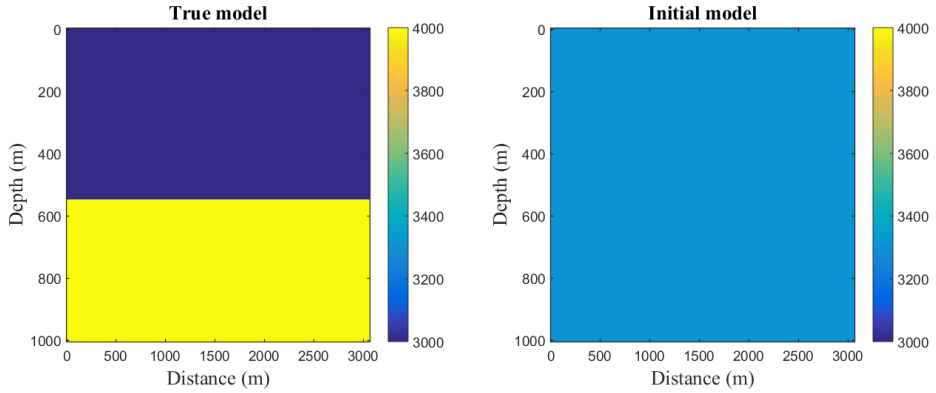
$\mathbf{p^n}$ is the pressure at time step n and $\mathbf{f^n}$ is the source term. The imaging condition used in time domain LSRTM algorithm is

$$\mathbf{m}_{mig}(\mathbf{x}) = \sum_{n_s}\sum_{n_t}\phi_\mathbf{s}(\mathbf{x},\mathbf{t})\phi_\mathbf{r}(\mathbf{x},\mathbf{T}-\mathbf{t}), \tag{47}$$

which represents the process of the source wavefield $\phi_\mathbf{s}(\mathbf{x},\mathbf{t})$ and the receiver wavefield $\phi_\mathbf{r}(\mathbf{x},\mathbf{T}-\mathbf{t})$ cross-correlation. This imaging condition is the same as the imaging condition in the frequency domain. In the following section, we will present some examples to compare time domain LSRTM and FWI-based LSRTM in frequency domain.

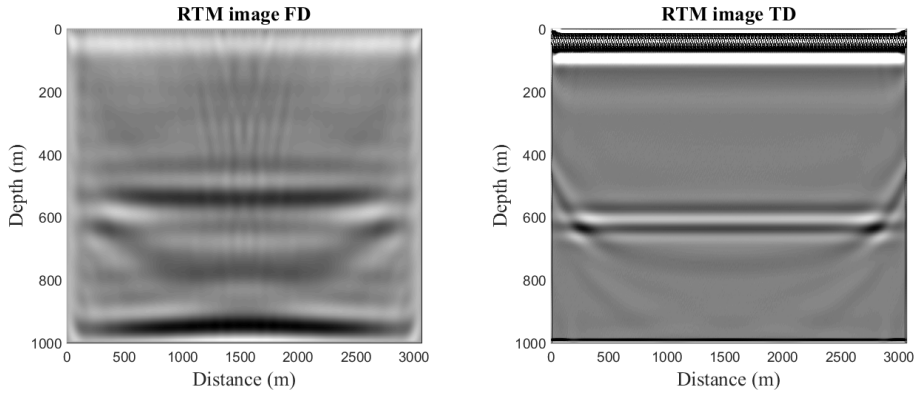
NUMERICAL EXAMPLES

In this section, we will show several tests to compare the time domain and the FWI-based frequency domain LSRTM for cases when the velocity model is very wrong. The goal is to observe if either approach is more robust than the other. The model dimensions for all the tests are 126 points for the vertical axis and 384 points for the horizontal axis and the grid spacing is 8m. FIG. 3. shows a 2-layer model with velocities of 3000m/s for the upper layer and 4000m/s for the lower layer. The interface is at depth 560m and the wrong velocity model is set to a constant value of 3300m/s. The migration results are shown in FIG. 4. Because the velocity of the first layer is wrong (3300m/s instead of 3000m/s), the reflector depth is wrongly mapped for the time domain RTM, to around 600m instead of 560m, deeper as expected for velocity too fast. On the other hand, the FWI-based frequency domain LSRTM seems to provide the correct location, although the resolution is poor. One of the goals of this ongoing research is to find out why the frequency domain migration has this lower resolution but more robustenss.



(a) The true model (b) The initial model

FIG. 3. The true model and the initial model

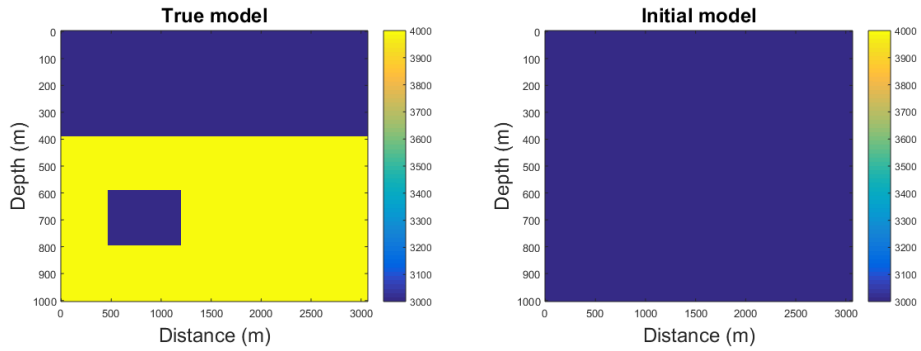


(a) RTM in frequency domain (b) RTM in time domain

FIG. 4. RTM of a 2-layer model in frequency and time domain

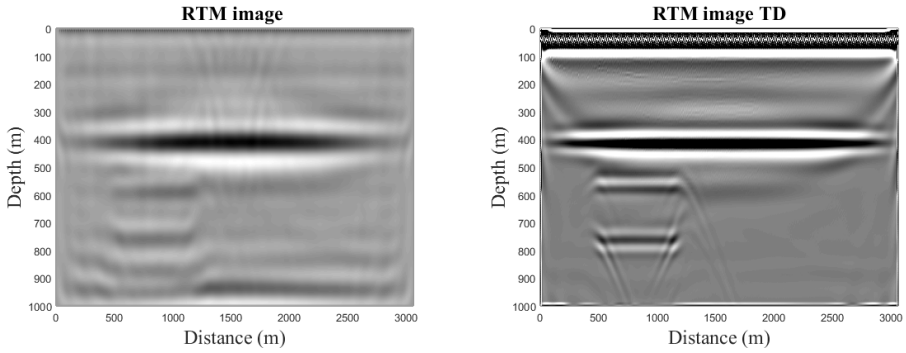
In the second test, we use a model with similar velocities as before, 3000m/s for the first layer and 4000m/s for the second layer, but this time we insert a small block in the second layer with the velocity of the first layer (3000m/s). We use a wrong velocity of

3000m/s, which is incorrect for the second layer. Therefore we expect that the reflector depth will be correct for both methods, but the block should appear smeared in both cases. Also, we expect that the LSRTM will not be able to improve focusing since residuals will not help convergence towards the true model (gradient will be wrong). FIG. 6. and FIG. 7. illustrate the results of RTM and LSRTM for this case. As expected, there is no obvious improvement from RTM when using LSRTM. The side boundaries are poorly imaged, which indicates poor focusing due to the wrong velocity. As before, the depth of the block seems more accurate in the FWI-based LSRTM. Because the reflector should have been correctly mapped for both cases, the fact that the frequency domain version shows low resolution indicates that this method suffers from convergence problems when the velocity is wrong, which does not affect the time domain version at the location of the first reflector. The nature of the frequency domain solution, direct instead of sequential as in the time domain, maybe the reason behind this observation.



(a) The true model (b) The initial model

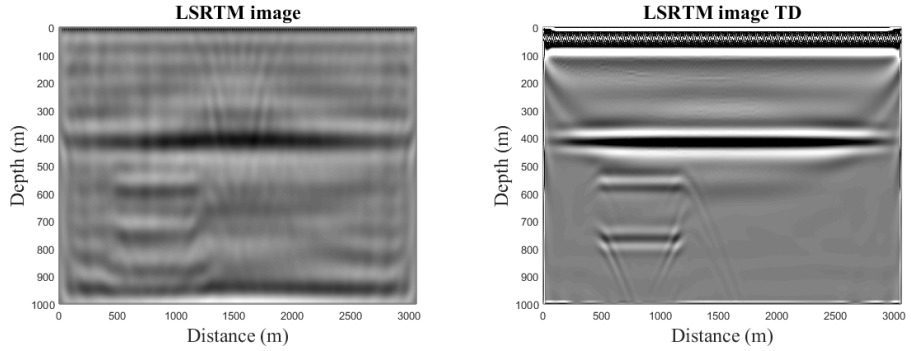
FIG. 5. The true model and the initial model for a blocky model



(a) RTM in frequency domain (b) RTM in time domain

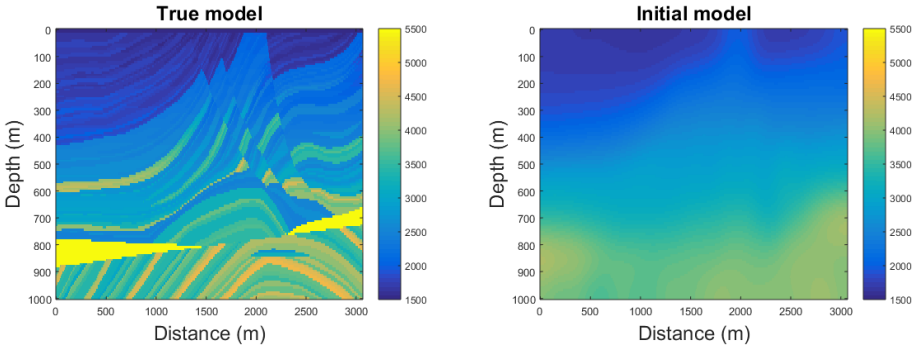
FIG. 6. RTM of the blocky model in frequency and time domain

In the final test, we use the Marmousi model, again with wrong velocities. To create the wrong velocities we apply a very heavy smoothing, much more than usually done to honour Born modeling approximation. FIG. 8. shows the true model and the smoothed model. Comparing the results in different domains (FIG. 9 and FIG. 10), the FWI-based LSRTM can locate the reflectors at approximately correct locations but suffers from the noise and convergence problems. The FWI-based LSRTM seems to have more details than the time domain version, in particular deeper in the section.



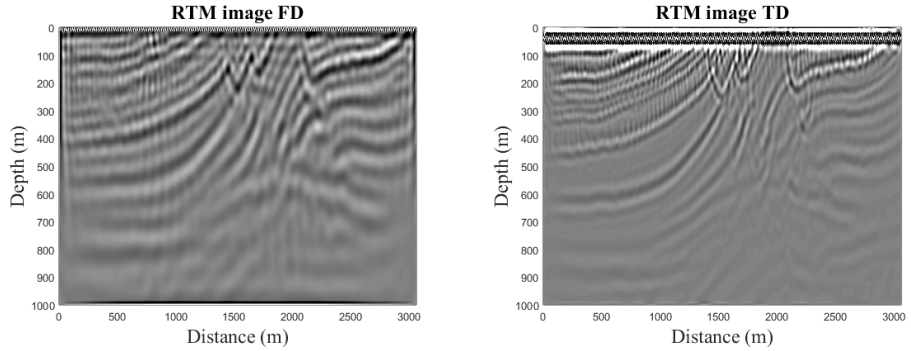
(a) LSRTM in frequency domain (b) LSRTM in time domain

FIG. 7. LSRTM of the blocky model in frequency and time domain



(a) The true model (b) The initial model

FIG. 8. The true model and the initial model of the resampled Marmousi model

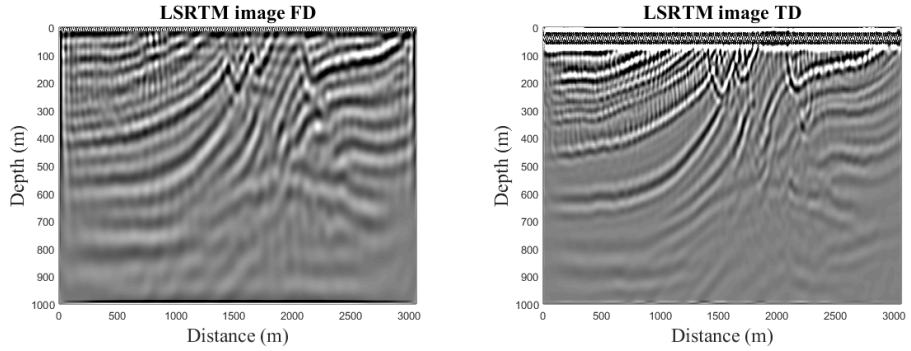


(a) RTM in frequency domain (b) RTM in time domain

FIG. 9. RTM of the Marmousi model in frequency and time domain

CONCLUSIONS

In this paper, we first compare the LSRTM in frequency and time domains and then propose a FWI-based LSRTM in the frequency domain. This method is based on the objective function of FWI and uses truncated Newton's method to solve for the model perturbation. We used three examples to compare these algorithms and understand the strength and weakness of each method. The conventional time domain LSRTM has a good convergence and sharper image when the velocity model is correct, but when the model is wrong, the reflectors are not correctly located. For the FWI-based LSRTM in the frequency domain



(a) LSRTM in frequency domain (b) LSRTM in time domain

FIG. 10. LSRTM of the Marmousi model in frequency and time domain

with wrong velocities, although it suffers from convergence problems, it seems to locate the reflectors at the correct locations. We speculate that low frequencies seem to be less affected by the wrong velocities. If this is true, we expect to be able to develop a RTM in the frequency domain that uses information from the low frequencies to constrain the high frequencies. Also, we plan to investigate further how the convergence is affected in each case when the velocities are wrong.

ACKNOWLEDGEMENTS

We thank the sponsors of CREWES for the support. This work was funded by CREWES and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 461179-13.

REFERENCES

Ajo-Franklin, J., 2005, Frequency-domain modeling techniques for the scalar wave equation: An introduction.

Chen, K., and Sacchi, M. D., 2018, Time-domain elastic gauss-newton full-waveform inversion via matrix-free adjoint-state method.

Dai, W., Fowler, P., and Schuster, G. T., 2012, Multi-source least-squares reverse time migration: Geophysical Prospecting, **60**, No. 4, 681–695.

Engquist, B., and Majda, A., 1977, Absorbing boundary conditions for numerical simulation of waves: Proceedings of the National Academy of Sciences, **74**, No. 5, 1765–1766.

Geng, Y., and Innanen, K. A., 2016, Frequency domain nonlinear full-waveform inversion: CREWES Research Report, **28**.

Geng, Y., Innanen, K. A., and Pan, W., 2017, Nonlinear multiparameter full waveform inversion based on truncated newton method: CREWES Research Report, **29**.

Métivier, L., Brossier, R., Virieux, J., and Operto, S., 2013, Full waveform inversion and the truncated newton method: SIAM Journal on Scientific Computing, **35**, No. 2, B401–B437.

Pan, W., Innanen, K. A., and Liao, W., 2017, Accelerating hessian-free gauss-newton full-waveform inversion via l-bfgs preconditioned conjugate-gradient algorithm: Geophysics, **82**, No. 2, R49–R64.

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