

Tutorial: the mechanics of waveform inversion

Ian F. Jones¹

Abstract

Unresolved velocity anomalies lead to distortion in images: consequently, much effort has gone into developing model-building techniques to identify such anomalies. Historically, the industry has relied on ray-based tomography to achieve this, but ray methods are limited to detecting features that are typically larger than about five times the dominant wavelength of the recorded seismic data. More recently, model building based on wavefield tomography has been introduced (full waveform inversion). Waveform inversion methods are more costly than ray methods, but have the potential to resolve features smaller than the recorded seismic wavelengths.

Using waveform inversion to update a parameter field comprises two main steps: firstly, determine the spatial location of where an observed error came from, and then, determine the magnitude of that error, so as to update the parameter model. The first step uses the same principles as reverse-time migration to construct an ‘image’ of the parameter error, and the second step employs gradient descent methods to estimate the magnitude of the required parameter update.

In this tutorial, I will describe both steps of the waveform inversion procedure, and also discuss differing methods of characterizing the error in a given parameter model.

Introduction

Migration can only construct reliable images of the subsurface if we have an accurate representation of the parameter variation within the earth. Consequently, our industry has spent many years developing model building tools to estimate these parameters, the most important of which is velocity. Velocity variation can be classified on the basis of the scale length of the variation as compared to the wavelength of the seismic wavelet. (Note that throughout this text, I’ll be using the word *velocity* very loosely: in general, I mean any parameter governing wave propagation). If the velocity scale length is much greater than the seismic wavelength, then ray-based tomography (using only travel-time information) can resolve the features. If not then this (high-frequency) ray approach is inappropriate, as diffraction (scattering) phenomena will predominate, and then waveform tomography (also referred to as ‘full waveform inversion’ (FWI) and ‘diffraction tomography’), which uses the wavelet-shape as well as arrival-time information, must be used instead. The techniques behind FWI were first introduced by Tarantola (1984; 1987; 1988) with other developments such as by: Lailly (1983); Worthington (1984); Mora (1988); Schuster (1993); Pratt et al. (1996; 1998); Sirgue and Pratt (2004); Plessix (2006, 2007); Symes (2008); Vigh and Starr (2008); Virieux and Operto (2009); Warner et al. (2010); Wang et al. (2011); Yingst et al. (2011); and Virieux et al. (2017). However, Tarantola’s ambition was to directly invert for physical earth properties rather than the somewhat less ambitious objective of simply making an improved model for depth migration.

Consider the situation with a velocity anomaly whose physical dimensions are much larger than the seismic wavelength. In this case, describing the propagating wave-front with representative ‘rays’ (normal to the wave-front) is acceptable as Snell’s law adequately describes the refractive and reflective behaviour at the interfaces of the anomalous velocity region. Conversely, once the velocity anomaly is of similar scale length to the seismic wavelet, then diffraction phenomena dominate, and it is then scattering which governs the behaviour of the wave-front. Consider a low-frequency sound wave encountering an irregularly-shaped boulder. As the sound-wave passed, the entire boulder would vibrate, radiating energy in all directions, effectively acting as a secondary source. Trying to apply Snell’s law at each point on the corrugated boulder’s surface would be meaningless.

Whereas travel-time tomography iterates with renditions of ray tracing, with waveform tomography one must iterate with renditions of the propagating waveform using repeated forward modelling with, for example, finite differences – which is more costly than ray-tracing (e.g. Pratt et al., 1996; 2002; Sirgue and Pratt, 2002; 2004; Plessix and Perkins 2009; Warner et al., 2010). Using the starting guess of the model, a finite difference modelling exercise (of limited bandwidth) is undertaken to make a synthetic version of the recorded field data. The real and synthetic modelled data are then compared (e.g. by subtracting), and the waveform tomography iterates to update the gridded velocity model so as to minimize the difference between the recorded and

¹ ION Geophysical

* Corresponding author, E-mail: ian.jones@iongeo.com

DOI: 10.3997/1365-2397.2019017

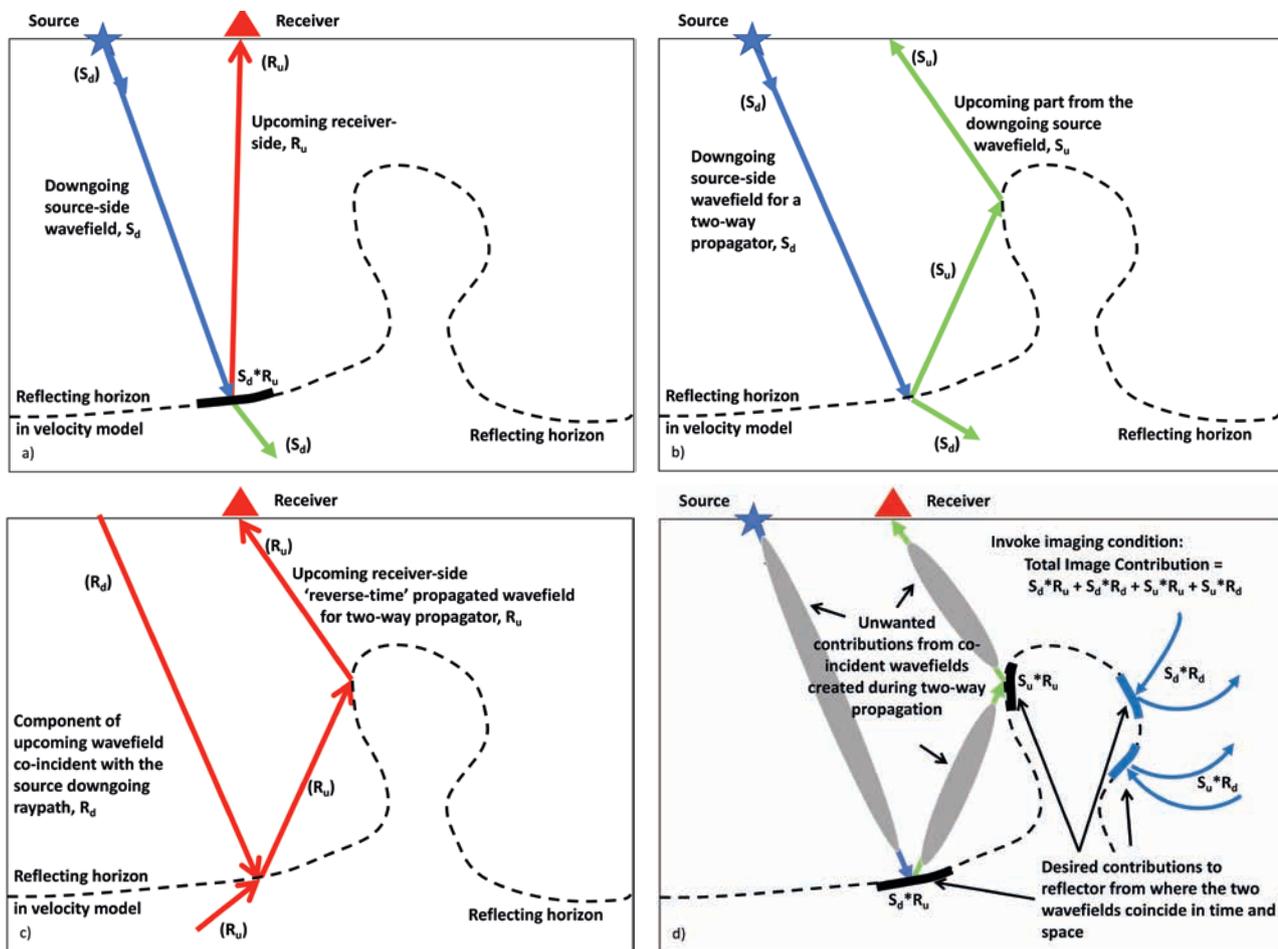


Figure 1 The migration imaging condition. a) Simple reflection from a flat event. b) Downgoing source-side wavefield for two-way propagation. c) Upcoming receiver-side wavefield. d) Imaging condition forms the final image plus cross-talk terms (e.g. the grey regions along coincident portions of the ray-paths). Adapted from Jones et al. (2017).

modelled data. In principle, this technique can resolve features smaller than the seismic wavelengths available in the recorded data, as real phase and amplitude changes are very sensitive to slight variations in the parameters governing propagation.

In this tutorial, I'll outline the key stages involved in constructing a waveform inversion algorithm (assuming that we have obtained a good estimate of the wavelet for use in the forward modelling), namely:

1. form a representation of model error,
2. determine where in the earth (which model cell) this observed error came from,
3. find the magnitude of the required velocity update for each model cell so as to minimize the model error.

In its simplest form, the first of these steps is achieved by simply subtracting the recorded field data and the forward modelled data (formed using the current velocity model). Later in the tutorial, I'll describe some other methods to achieve this.

The second and third tasks are much more demanding, and I'll begin this tutorial describing them. However, before delving into the vagaries of waveform inversion itself, I'll first review and recap the details of subsurface imaging based on wave propagation. The reason for this diversion is that the principles employed in wavefield extrapolation migration are also central to the second

step in FWI (namely determining *where* in the subsurface an observed error originated).

Determining *where* something came from: using wavefields to build images

Building an image based on wave propagation, relies on forward modelling of waves emanating from the source location, and reverse modelling of the waves recorded at the actual geophones or hydrophones. These two modelling exercises use a full solution of the acoustic wave equation, in order to handle energy travelling in any direction (the two-way wave equation), and the industrial standard algorithm for imaging with this method is reverse-time migration (RTM).

How can we describe a reflection in terms of the downgoing and upcoming wavefields? By definition, if the downgoing and upcoming wavefields are in the *same place* at the *same time*, then there must be a reflector there. This is Claerbout's imaging condition principle (Claerbout, 1971; 1985). Employing a finite-difference (FD) algorithm, we can use this principle to form an image by:

1. Forward modelling a synthetic shot (i.e. evaluating what the expanding wavefront looks like at incremental steps of a few milliseconds of propagation time).
2. Back-propagating the recorded data (i.e. evaluating what the recorded field data looked like for previous propagation

times, working backwards from the final recording time to time zero).

3. Multiplying these two extrapolated wavefields together at each corresponding propagation time-step.
4. Summing the results for all propagation steps.

As the above procedure involves summing products, it constitutes a convolution, and is thus referred to as the convolutional imaging condition in shot migration: the image is being formed by what is essentially a convolution of downgoing and upcoming wavefields (e.g. Bancroft, 1997, 2007). This process is repeated for all available shots, and all these overlapping 3D shot-contribution volumes are summed to form the full migrated image of the study area. A tutorial overview of these concepts is given in Jones (2014) and Jones et al. (2017).

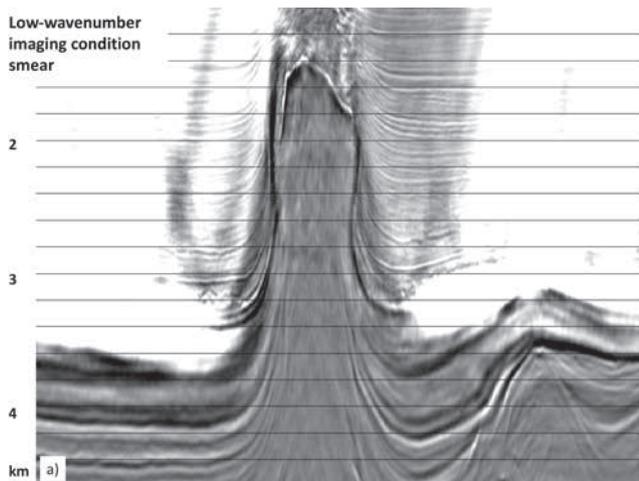
Cross-talk imaging artefacts in RTM: ‘one man’s meat is another man’s poison’

In its simplest form, the convolutional imaging condition for two-way acoustic migration produces four image contributions, not all of which may be desirable at any given subsurface location. Consider a source wavefield S , with downgoing and upcoming components S_d and S_u , respectively and a back-propagated receiver wavefield R , with upcoming and downgoing components R_u and R_d . The convolutional imaging condition for a shot record is formed by the superposition of the source and receiver wavefield products at all coincident times integrated over time:

$$\text{Image} = S * R = (S_d + S_u) * (R_d + R_u) = S_d * R_u + S_d * R_d + S_u * R_u + S_u * R_d \quad (1)$$

This is followed by summation of all shot contributions (e.g., Liu et al., 2011).

For a simple reflection from one-way propagating wavefields illuminating a flat reflector, only the first of these terms $S_d * R_u$ would represent the desired physical image (Figure 1a). If using a two-way propagator (Figure 1b, c) then an image of a steep event illuminated by a double bounce on the receiver-side would be formed from the terms $S_d * R_d$ and by a double bounce on the source-side from the terms $S_u * R_u$ (Figure 1d).



Alternatively, if we had an overhanging interface in a medium with significant vertical velocity gradients where both the source and receiver-side wavefields turned, then the image could be formed from these turning waves reflecting from the underside of the overhang: in this case the useful contribution would be formed from the terms $(S_u * R_d)$. Hence, RTM has the benefit of being able to form an image from illumination along complex travel paths, but has the downside of producing all combinations of the wavefields even when they do not physically exist, i.e. the crosstalk noise problem. The low frequency low wavenumber background ‘image condition’ noise is seen in most places in a raw RTM image above the point where the critical angle is reached for either the downgoing or upcoming wavefields. These cross-talk artefacts will be revisited a bit later, as it turns-out that they constitute the ‘rabbit ear’ components of the FWI gradient, and are of use in updating the parameter model.

Figure 2 shows a real salt diapir ‘raw’ RTM image before and after filtering to remove cross-talk contamination. In Figure 2a, the low frequency background ‘image condition’ noise is seen in most places above the point where the critical angle is reached for either the downgoing or upcoming wavefields. After suppression of the noise with filtering of the image or angle gathers, we obtain the result shown in Figure 2b.

Using the RTM imaging condition to facilitate model update

Now consider a twist on the approach just described for imaging using RTM, and rather than asking where in the subsurface a reflection came from, let’s try to ascertain where in the subsurface an observed velocity error came from. First let’s state what we’d like to achieve in order to solve the problem: iteratively modify the parameter model so as to *minimize the difference* between observed waveform amplitudes on the real data, and modelled amplitudes in synthetic shot gathers created using the current parameter model.

As with RTM, we use an imaging condition to combine back-propagated receiver-side information with forward-propagated source-side information, to tell us ‘where in the subsurface’ something came from. With RTM we seek the spatial

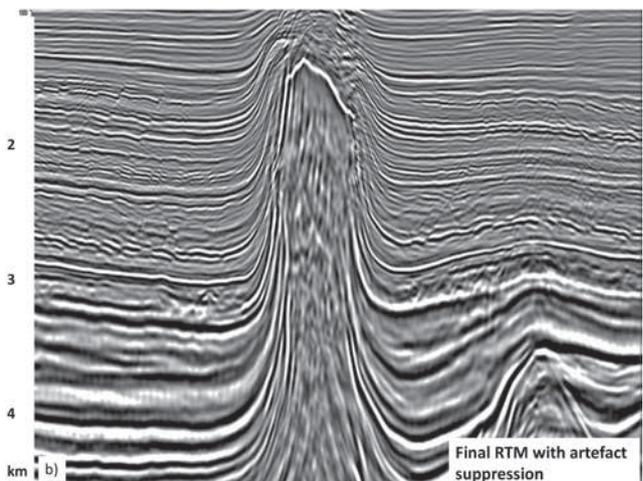


Figure 2 a) Shallow section from the RTM image prior to filtering the backscattered noise; b) Image from later stage in the model-building after filtering of RTM angle gathers (From Jones and Davison, 2014: ION RTM image shown courtesy of Talisman Sinopec Energy UK and partners GdF-Suez, EON, & Idemitsu. Input data courtesy of CGG).

The Gradient: FWI procedure for a single shot

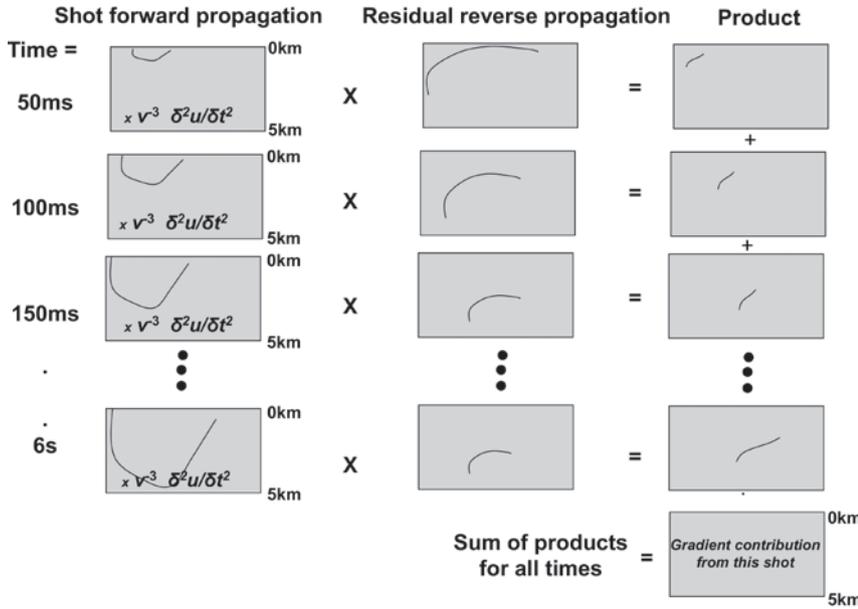


Figure 3 Schematic of the formation of the gradient: the second time derivative of the forward propagation source wavefield is ‘convolved’ with the back-propagated residual.

location of an image contribution from a given shot gather, but for FWI we seek the *spatial location of the velocity model error* that produced the observed shot gather ‘residual’ (i.e. a characterization of the difference between the observed and modelled shot record): this was described by Tarantola (1984). In its simplest form, the residual will be the difference between the observed and the modelled data, but could also be formulated as the shaping filter that matches them (e.g. Warner and Guasch, 2016) or the time or phase shift that aligns them (e.g. Vigh et al., 2016; Vigh et al., 2017; Fu et al., 2017; Schuster, 2017). As well as determining *where* the error came from, we also need to formulate a relationship between the magnitude of the error, and the actual parameter perturbation at that location that will ‘fix it’.

Given that we want to minimize this ‘residual’ (often simply the difference between the observed and forward modelled data), how do we set-up the problem to achieve this numerically? We can use least squares minimization. For observed data d_o and synthetic data $d(m)$ created with model m , the residual is:

$$residual(m) = [d_o - d(m)] \tag{2}$$

In its simplest form, the ‘cost function’ $C(m)$ (also known as the ‘objective function’) that we need to minimize is:

$$C(m) = \sum \sum \sum [residual(m)]^2 = \sum \sum \sum [d_o - d(m)]^2 = \langle d_o - d(m) \rangle^2 \tag{3}$$

Summing over: all shots in the survey, all traces in the shot-gathers, all time samples in the traces.

In order to minimize the cost function (evaluate its variables when it approaches zero), we form the derivative of $C(m)$ with respect to the model parameters (in most cases, velocity) to relate the change in $C(m)$ to change in the model:

$$\begin{aligned} \partial C(m) / \partial m &= \partial [\langle d_o - d(m) \rangle^2] / \partial m \\ &= -2 \langle d_o - d(m) \rangle \partial d(m) / \partial m \end{aligned} \tag{4}$$

Note: the field data, d_o is *NOT* a function of the model we are using.

This derivative will ideally become *zero* when we have the ‘correct’ model values: i.e. for model $(m+\Delta m)$, where Δm is the perturbation required to correctly update the initial model m . So, writing: $\partial C(m+\Delta m) / \partial m = 0$, and expanding this as a Taylor series to 1st order gives:

$$\partial C(m+\Delta m) / \partial m = \partial C(m) / \partial m + \partial^2 C(m) / \partial m^2 \cdot \Delta m = 0 \tag{5}$$

Or, after rearranging, we note that the desired model perturbation is:

$$\Delta m = - [\partial^2 C(m) / \partial m^2]^{-1} \cdot \partial C(m) / \partial m \tag{6}$$

where the term $\partial C(m) / \partial m$ is referred to as the gradient (i.e. the ‘direction’ of required model parameter update), and $[\partial^2 C(m) / \partial m^2]$ is referred to as the Hessian.

To determine Δm , we thus need to solve for the inverse of the Hessian, but this is either costly or impractical, so this is typically addressed with an approximate iterative inverse technique. We also need to evaluate the gradient (the derivative of the cost function: $\partial C(m) / \partial m$). What this entails is finding the spatial location of the current observed residual error: we can achieve this by using the RTM engine to back-propagate this error in conjunction with an estimate of how the downgoing source side information also changes with respect to model error. This latter aspect is evaluated as follows: recall that the propagating wavefront is governed by the (acoustic) wave equation:

$$\nabla^2 u = v^2 \partial^2 u / \partial t^2 \tag{7}$$

where v is the P-wave propagation velocity, and u is the particle displacement. Consequently, to consider how the behaviour of this propagating wavefield will change as a function of altering

the model parameters (say, velocity, v), we can take the derivative of Equation 7 with respect to v :

$$\partial/\partial v (\nabla^2 u) = -2v^{-3} \partial^2 u / \partial t^2 \tag{8}$$

Then instead of forming the usual RTM image as the convolution of the back propagated data and forward propagated source wavelet, we can form a new object as the convolution of the back propagated residual with the second time-derivative of the forward propagated source term scaled by $-2v^{-3}$. In this scheme, the residual is referred to as the ‘adjoint source’ (e.g. Warner and Guasch, 2016).

$$\text{Gradient} = \partial C(m)/\partial m \propto \int v^{-3} \partial^2 [\text{forward-propagated source}] / \partial t^2 \cdot [\text{back-propagated residual}] \tag{9}$$

We can use the RTM engine in this way so as to iteratively converge on the required velocity update, by first localizing the spatial locations of the underlying parameter errors, and then using a gradient descent scheme to estimate their numerical values. This first part of the procedure is summarized in Figure 3.

In the time-domain, evaluation of Equation 9 requires keeping track of the second derivative of the down-going source wavefield with respect to time during the downward extrapolation procedure, which can be achieved by computing and storing these estimates during the forward propagation of the source-side term (and remember that the input source term represents the wavefield at its starting point $t=0$). So to achieve this, we require a reasonable estimate of the source wavelet, which can be non-trivial, especially for land data where the downgoing wavefield may change rapidly from shot-to-shot due to variation in near surface and coupling conditions.

As noted, the first step in this process is to ascertain *where* in the 3D earth-model, the velocity error came from, with a notion of whether we need to increase or decrease the velocity i.e. the ‘direction’ of change, or ‘*gradient*’. The gradient, as defined in Equation 9, is obtained via an imaging condition using the back-propagated residual and the second time derivative of the forward propagated scaled source term, integrated over all propagation times.

This gradient gives us an initial estimate of the direction of the required velocity change, i.e., whether we need to increase or decrease parameter values. In other words, we’ve converted the pseudo image change into an associated velocity change. However, this is only a crude initial estimate of the *direction* of the required update: as usual, many approximations were involved. Also, this gradient is essentially only one snapshot of the associated model error relative to the current model realization, and in order to update the model sensibly, we need to evaluate the cost function for a range of model parameter values, so as to move towards an overall (‘global’) minimum.

Consequently, once we have the general direction of the update (the raw gradient), we then need to ‘fine-tune’ the amount of update required for this gradient direction, so as to get closer to the minimum in the cost function: this is where terms such as ‘line search’, ‘step length’, ‘steepest descent’, etc. are heard.

Finding the minimum in the cost function

Given the raw gradient formed as mentioned, we now loop over several iterations of changing the model parameters, repeating the forward modelling to create new synthetic shots, reforming the residual, recomputing the cost-function, and checking to see if the cost function is getting smaller. After several forward modelling exercises with the perturbed model parameters, having converged to an acceptable reduction in the cost function for the current gradient direction (i.e. several iterative steps along the line search direction), we start all over again, once more using the RTM engine to back-propagate the latest residual and forward modelling the (scaled) source term, to obtain a new ‘raw’ gradient. Using this new raw gradient direction, we recommence the line search. This procedure continues until we find a satisfactory (and hopefully global) minimum in the cost function.

This procedure for finding the (global) minimum in our objective function can be addressed with many numerical methods (steepest descent, non-linear conjugate gradient, etc.). I’ll show some graphical representation of this procedure first for one, and then for two variables (model parameters).

Consider the velocity model in Figure 4a, where we are happy with the values of velocity (and/or all other parameters) in all cells *except* for one. If we perturb the value of this ‘bad’ velocity,

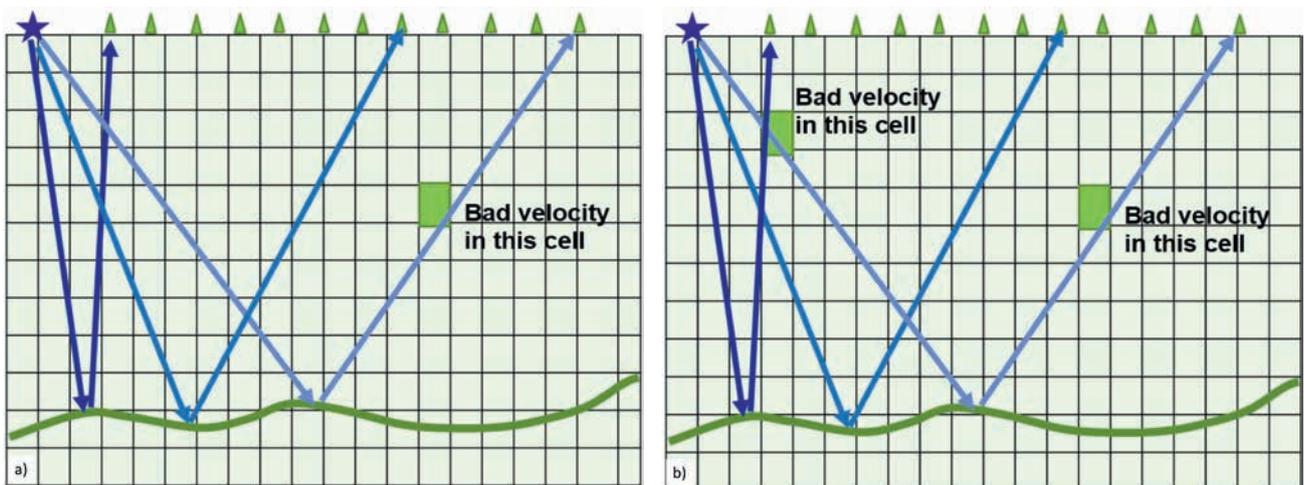


Figure 4 a) Gridded velocity model with just one bad value of velocity. b) Gridded velocity model with two bad values of velocity.

we can observe its effects on the synthetic data computed with each of the single-cell perturbed models. If we look just at say a single (far offset) trace with one reflection event, and see what happens to the measurable arrival time as it shifts slightly due to the velocity perturbations we impose on the single ‘bad’ cell, then we can form the residual trace for each of these shifts (Figure 5a) and from this compute the cost-function given by Equation 3. In its simplest form the cost function is just the sum-of-squares of the residual for all times, all traces, and all shots, often with some muting of undesired events (as will be described later, it can be beneficial to deal with reflected and refracted energy separately, sometimes also isolating just the very first arrival refraction events). The cost function $C(m)$ for this residual is plotted in Figure 5b.

The cost-function shown in Figure 5b was for a single waveform, and for the variation in a single parameter (the velocity giving rise to the far-offset arrival-time), so the cost function could be plotted on a single axis. Now consider the velocity model shown in Figure 4b, where we have two ‘bad’ cells. If we jointly varied these two parameters, we would need two axes to plot the cost function, and could then plot it as a contour plot in a 2D plane.

Figure 6 shows this 2D plane in ‘model space’ for the two variables we are perturbing (the two bad velocity values). We could plot the cost function $C(m)$ along one of these axes for change in a single variable, as was done in Figure 5b. Alternatively, we could evaluate $C(m)$ for all possible combinations of the two model parameters to produce a contour surface of $C(m)$, as shown in Figure 6b. (In order to simplify the figures, I have used a simple bowl-shaped surface to represent the central region of the cost function $C(m)$, rather than the oscillatory function shown in Figure 5b).

However, this is difficult to visualize for thousands of parameters in a real problem (all the cells in the 3D model), and rather than the cost function existing in a 2D plane, it exists in what is sometimes referred to as a hyper-plane (or more correctly, a higher dimension Euclidean space). Remember that in a realistic

3D velocity model spanning, say 1000 km² to a depth of 10 km, with model cells of dimensions 100 m*100 m*100 m, we will have 10 million cells, and each cell might have more than one parameter ($v_p, \epsilon, \delta, \rho, Q, \dots$), all of which might require an update.

For example, consider a model whose starting velocity values were at the slownesses represented by the blue dot in Figure 6a, with the cost function’s global minimum at the green dot. Ideally, to solve Equation 6 in order to find the desired Δm , we would have to evaluate $C(m)$ for all possible combinations of the two parameters in order to produce the contour values of the cost-function’s surface, so as to identify the location of the overall minimum (the green dot). But for the millions of parameters in a realistic problem, this is impractical. Consequently, we try to solve the problem approximately and iteratively. First, compute the gradient at the starting location (blue dot) by using the pseudo imaging condition of Equation 9, and extract $C(m)$ along the line indicated by the gradient direction at this location (blue dashes). The values of $C(m)$ along this direction produce a curve in that 2D plane, and this curve will have a minimum somewhere, as indicated by the yellow dot on the perspective view (6b). If we now step along the $C(m)$ curve in the gradient direction to get to this minimum (yellow dot in 6c), this will represent a new model, which will be a bit closer to the global minimum (the green dot).

However, given that we do not inherently know the required length of the step in the current gradient direction, we need to search along this line segment in smaller steps until we find the current local minimum for this gradient direction (Figure 6d). At each one of these sub-steps, which corresponds to a new set of model parameters, we have to recompute the forward modelled shots, recalculate the residual and associated cost function, and see if the new cost function is acceptably small(er) than the initial cost function. Note that this really only gets us closer to the minimum for this gradient direction, as there is no guarantee that a step will bring us right to the actual (yellow dot) minimum. Once we have reached the local minimum for this particular stepping direction, we start over, and search for a new direction of the downhill gradient, once again using the RTM engine, and step

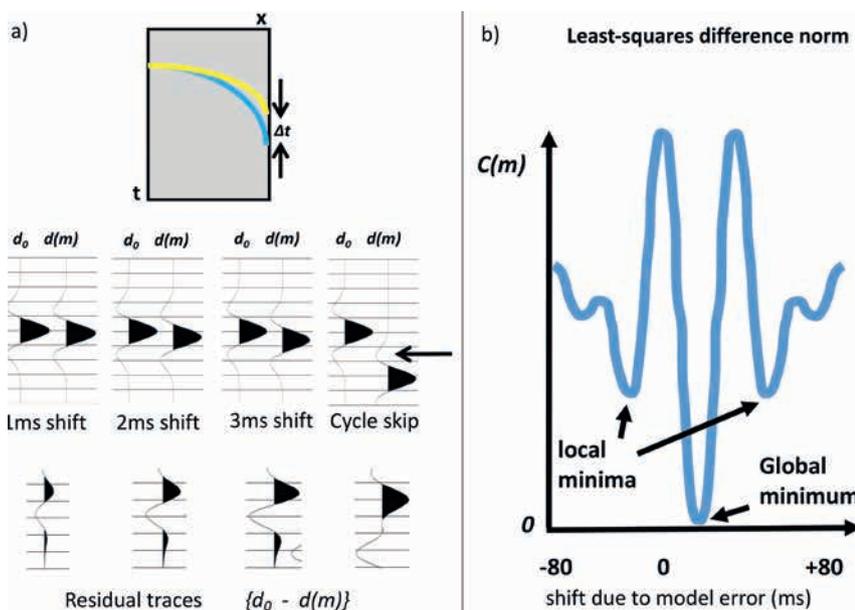


Figure 5 a) Top: time shift Δt between the field data (yellow) and modelled data (blue). Centre: far offset traces from ‘field’ and modelled data for various model errors. Bottom: residual traces. b) For an oscillatory waveform (50Hz Ricker wavelet) the cost function for various time shifts resulting from the velocity error can be computed as the sum of squares of the residual.

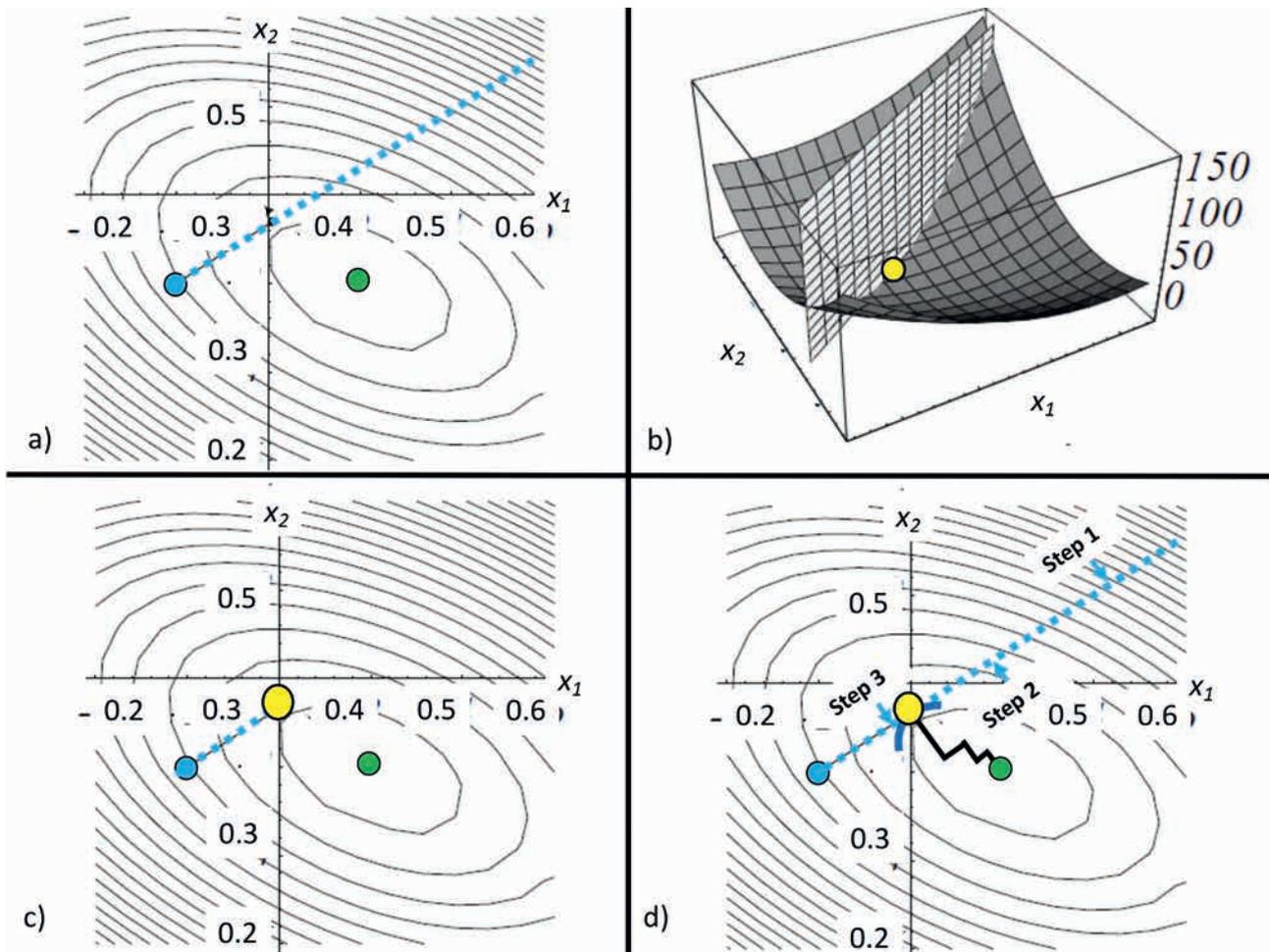


Figure 6 a) Contour plot of $C(m)$ for two variables, indicating the starting point for model building (blue dot) and its gradient direction (dashed blue line). b) $C(m)$ curve in 2D plane along gradient direction with minimum at the yellow dot. c) Contour plot showing minimum in the line-search direction (the yellow dot), before then constructing a new gradient to start over in a new direction. d) For a given gradient direction we search in small steps to locate the local-line-minimum (the yellow dot), before then constructing a new gradient to start over in a new direction. In these graphics, the axes represent slownesses in s/km. Adapted from Shewchuk (1994).

off in that line-search direction. If the search direction follows the maximum downhill dip (i.e. at right angles to the contour), then this is called steepest descent method. At the minimum for this particular line-search direction (at the yellow dot), the gradient direction is tangential to the $C(m)$ contours (as indicated by the solid blue curve segment), and thus the next gradient computation (the next iteration of FWI) will yield a new gradient direction at right angles to the contours. After many right-angled changes in direction, we hope to get closer to the global minimum (green dot).

Consequently, each ‘iteration’ of the FWI procedure is made-up of two parts: firstly using the RTM engine to compute the current gradient direction, and then the subsequent search for an acceptable minimum in the current line search direction. Repeating this line-search stepping procedure along the current gradient direction, and then for all subsequent new gradient directions, will eventually (hopefully) get us somewhere closer to the minimum (green dot in Figure 6d): N.B. the point where the minimum is reached for a given line search direction, is tangential to the contours. This particular technique is called the steepest descent method, but this is not usually used in industrial FWI algorithms as its convergence rate is inferior to methods such as the non-linear conjugate gradient approach (Shewchuck, 1994).

The components of the gradient

There are several contributing parts to the overall gradient, related to the various possible ‘wave-paths’. Let us consider the origin of these paths, and their separation (as it turns out that it can be beneficial to deal with them independently).

In the conventional convolutional imaging condition in RTM, the cross-talk terms are undesired, and constitute a class of noise that we typically remove with a Laplacian filter. Recall from Equation 1 that the two-way migration imaging condition produces four terms involving the up and down components of both the source and receiver wavefields. In Figure 1d, even though we do not want the ‘grey’ cross-talk terms for the image, the downgoing source energy is certainly ‘sensitive’ to the velocity regime through which it passes, as is the ‘incorrectly’ downgoing non-physical back-propagated receiver-side term. So the non-physical parts of the propagation exercise are of use in assessing the behaviour of the current model vis-à-vis the observed field data.

These coincident non-physical cross-talk wave paths give rise to what is sometimes referred to as the ‘sensitivity kernel’ and also as bunny or rabbit ears, and can be of use in helping to update the velocity model. With these notions in mind, we can decompose the overall response of the back-propagated residual

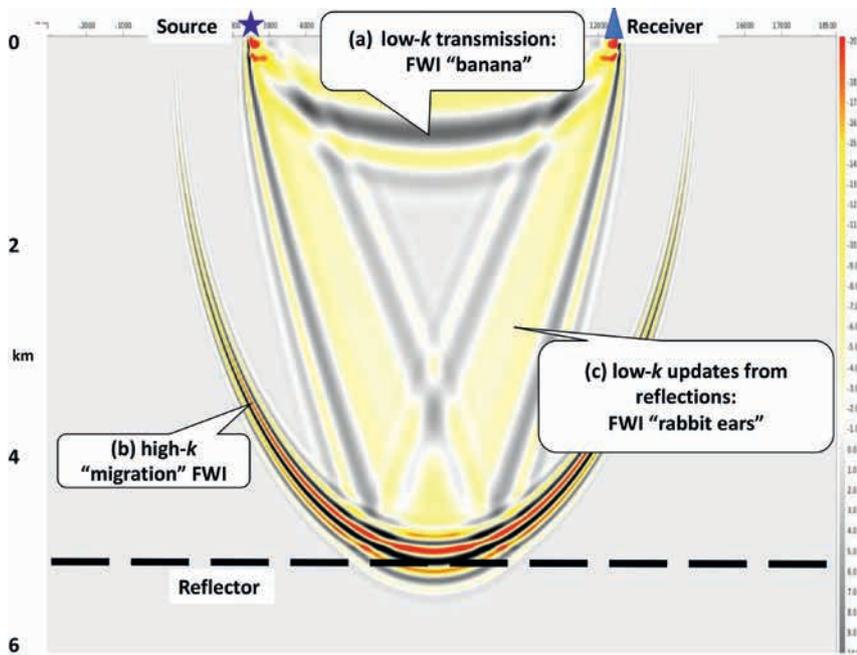


Figure 7 The overall gradient response is comprised of three constituent parts, which are the direct arrival ‘banana’ refraction path (a), the usual migration-response-like elliptical arc (b), and finally the ‘cross-talk’ rabbit ears which have much lower amplitude (c). (Courtesy of Chao Wang, ION)

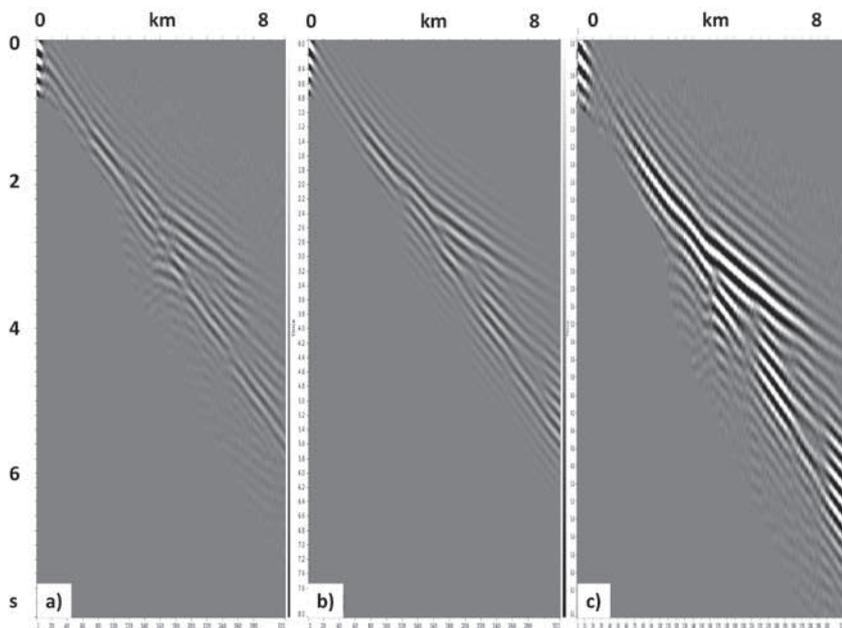


Figure 8 a) Single ‘field’ shot record, b) The modelled shot for the initial model, c) The residual shot before FWI. The field and modelled shot records should be balanced to give an overall amplitude match. Synthetic data courtesy of Chevron.

(the gradient), into three constituent parts, which are: the usual migration-response-like elliptical arc; the direct arrival ‘banana-shaped’ transmitted wavefield (refraction) path; and finally the ‘cross-talk’ rabbit ears (Figure 7). It can be shown that working independently with these separated gradient components can facilitate faster and more reliable model update (e.g. Bevc et al., 2008; Ramos-Martinez et al., 2016; Vigh et al., 2017).

For conventional RTM imaging, we would mute-off the refracted and direct arrival energy, and then use various filtering techniques to remove the cross-talk noise from the final RTM image and gathers. However, as can be noted in Figure 7, these cross-talk (‘bunny ear’) terms do penetrate deep into the model, reaching well below the direct wave (‘banana’ shape response), and importantly, their shape will be dependent on the velocity model being used in the migration. Hence, we can exploit them in the FWI model update scheme, and they can be vital in

facilitating a deeper model update than is offered by the refracted wavefield alone. The relationship between noise filtering in RTM and use of these terms for model update was discussed by Douma et al. (2010).

The form of the residual

In the technique described so far, the ‘residual’ was formed from the difference between the field data shot records and the synthetic forward modelled shots. However, this simple procedure is prone to producing cycle skipping (Warner et al., 2013) which is likely to cause the FWI to fail (i.e. to get stuck in a local minimum in the parameter space). This can be noted from Figure 5a: in the fourth model perturbation (shown on the right of the centre panel), the modelled wavelet is labelled as being cycle skipped, as the trailing trough of the ‘field data’ is aligned with the leading trough of the ‘modelled data’ (indicated with an arrow), and the

residual $C(m)$ can then be smaller than it would be for other intermediate alignments when no peaks or troughs were aligned.

What can we do to mitigate this problem? We could perhaps firstly put in a lot of effort with another model building technique, such as ray-tomography and hope that it was a good starting point for FWI (i.e. close enough to the global minimum), or acquire data with very low frequencies which suffer less from cycle skipping (<2 Hz), or we could form alternative ‘residuals’ such as those based on optimal transport schemes (e.g. Yang and Engquist, 2018) or the better known alternatives listed below.

- Design a matching filter that makes the real data look like the synthetic data (or vice versa). Iterate until this filter becomes a ‘spike’ (a band limited delta function). Once the filter is ‘spike-like’ it must mean that no more matching is needed as the real and modelled shots now resemble each other closely. In other words, we have converged on an acceptable velocity model (e.g. the adaptive-FWI method of Warner and Guasch, 2014a, 2014b; 2015).
- Correlate the real and synthetic traces (in sliding windows) to compute a time-shift or associated phase-shift, and iterate until the time or phase shifts become zero (this approach begins to look like ray tomography, as we are only considering the kinematics). Once the time-shift approaches zero, then

the real and modelled data must again closely resemble each other. In other words, we have converged on an acceptable velocity model (e.g. the adjustive-FWI approach of Jiao et al., 2015; Vigh et al., 2017; Wang et al., 2018). There are also variants of this approach that perform the time-shift matching only for a few selected arrivals (a ‘skeletonized’ approach, e.g. Schuster 2017): this then looks even more like ray tomography, as it involves picking and using just a few events. If just the first arrival refractions are selected, then this FWI variant can be thought of as a ‘fat-ray’ refraction tomography (e.g. Zhang, 1999).

- Concentrate only on the phase behaviour of the observed and modelled data, e.g. by replacing the amplitude spectrum of the modelled data with that of the field data trace-by-trace in small sliding windows (e.g. Maharramov et al., 2017; Schuster et al., 2017).

The appearance of the gradient

In order to see how an individual shot contributes to the overall update, and for purposes of instruction, let us first back propagate a single wavelet from within a shot’s residual, and then the entire shot, and look at their contributions to the gradient: this will produce something analogous to a migration impulse response. It

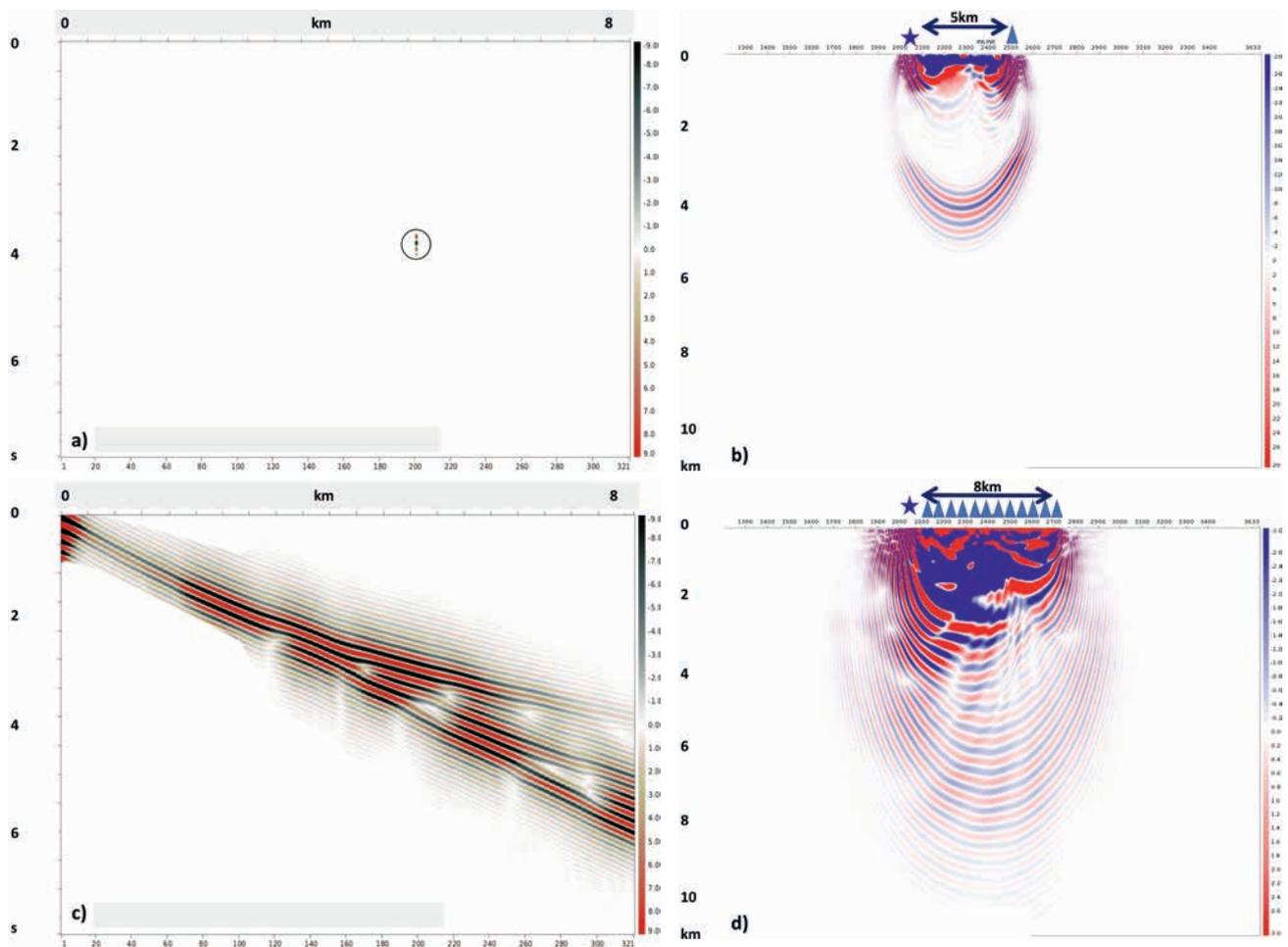


Figure 9 a) Residual gather with just a single wavelet, at offset 5 km and around time 3.7s (shown in the black circle). b) Gradient contribution from the single wavelet. The locations of the source (blue star) and receiver at 5 km offset (green triangle) are indicated. c) Entire residual shot gather and its gradient contribution d). Synthetic data courtesy of Chevron

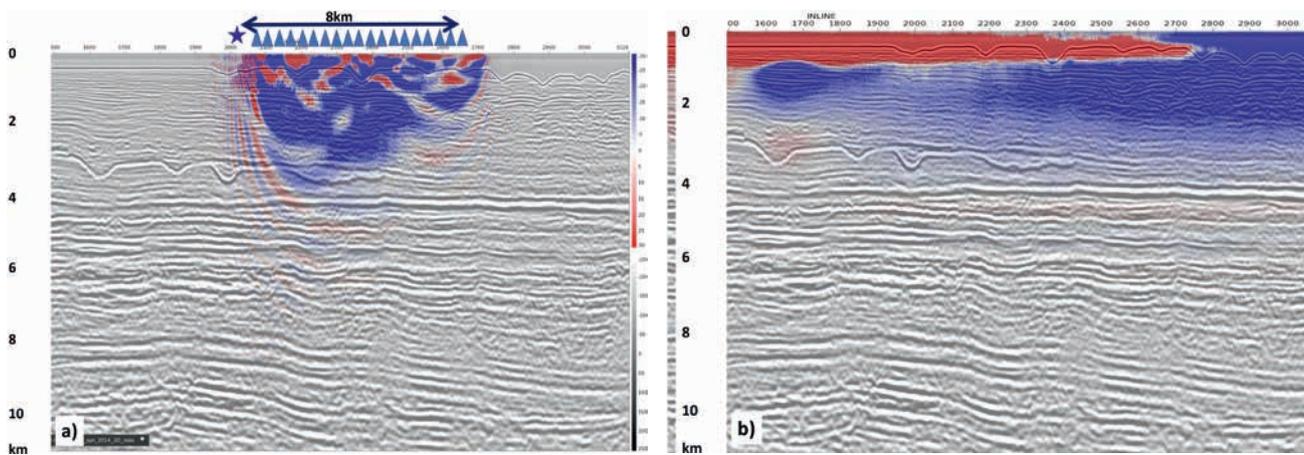


Figure 10 a) PreSDM image with the single shot's gradient contribution superimposed. The source location and the 8 km cable length are indicated. After summing the gradient contributions for all shots, we see the resulting final gradient in b). Blue indicates that we need to increase velocity, and red indicates a required decrease.

tells us that the model update contribution could have come from *anywhere* along the observed impulse response shape; we need to sum the gradient responses for many traces and many shots so as to rely on destructive and constructive interference to build a meaningful error contribution location (as indicated by the total gradient).

Figure 8 shows a single field shot record from 2D synthetic data, the modelled shot corresponding to the starting model, and the residual gather, before FWI. Deleting all values in the residual gather, except for a single wavelet, at offset 5 km and around 3.7 seconds, yields the gather shown in Figure 9a. The surviving wavelet is shown within the black circle. Subjecting this residual gather to the FWI procedure produces the gradient contribution shown in Figure 9b. The locations of the source (green star) and receiver at 5 km offset (green triangle) are indicated. Repeating this procedure for the entire residual shot gather (as is usually done, albeit with some judicious muting) produces the gradient contribution shown in Figure 9d. It can be seen for all these gradient images that they closely resemble migration impulse responses, and include typical migration 'noise'. Comparing the gradient elements from Figure 9, it can be noted how they resemble the elements of the simple synthetic gradients in Figure 7.

Figure 10a shows the preSDM image for these 2D synthetic data with the single shot's gradient contribution (from Figure 9d) superimposed. The source location and the 8 km cable length are indicated. After summing the gradient contributions for all shots from the seismic survey, we see the resulting final gradient in Figure 10b: this is now something that is useful and indicates where we need to increase or decrease the velocity in the model. However, given that the gradient is constructed using a migration procedure, it can suffer from migration noise and acquisition footprints as with imaged data. Hence, data conditioning of the gradient at each iteration can be beneficial, especially for narrow azimuth data where the acquisition footprints will be more pronounced. Note also that it is unlikely that there will be meaningful contributions to the gradient from within the water layer, hence we need to mute the gradient contribution in the water layer, as this contribution essentially constitutes migration noise.

Discussion: practical aspects of industrial implementation

During FWI model update, we typically iterate over many dozens of gradient computations, each with a few line-search steps. In these iterations, we can adjust the frequency and offset ranges used. The cost function shown in Figure 5 for a broad bandwidth wavelet was highly oscillatory, producing various local minima associated with cycle skipping between the field data and modelled data waveforms. For lower frequencies, the cost function will be less oscillatory, and thus cycle skipping is less likely. Consequently, for the earlier iterations of FWI it is beneficial to concentrate on the lowest available frequencies in the data. For deep subsalt model update, frequencies below about 2 Hz would be necessary to facilitate model update (e.g. Shen et al., 2017). Also, given that the cost of RTM is proportional to the fourth power of the maximum frequency (e.g. Farmer et al., 2006), iterating with lower frequencies is cost-effective.

When using the acoustic approximation, we ignore density changes, and consequently the modelled reflection amplitudes will be in error. Hence for the early stages of model update we tend to rely on the diving-wave energy (i.e., the transmitted or refracted wavefield). Unfortunately, this only tends to penetrate the subsurface to a depth of about one third of the maximum source-receiver offset, and if there are significant near-surface velocity inversions, we may not receive any useful refracted energy (e.g. Jones, 2018).

Hence, the state of the art (at least for data domain waveform inversion) tends to rely on long offsets, low frequencies, and to a large degree, turning wave data (Plessix and Perkins, 2009; Vigh et al., 2009; Sirgue et al., 2009; 2010; Plessix et al., 2010; Wang et al., 2011; Sirgue et al., 2011; Shah et al., 2012).

Once the refracted wavefield has been exploited to update the shallow sediment model, the reflected wavefield can be used. Various techniques exist to do this, primarily using the 'rabbit ears' resulting from the RTM cross-talk terms shown in Figure 7 (e.g. Ramos-Martinez et al. 2016; Vigh et al., 2017), or more recently, exploiting what is referred to as an 'extended condition' in the FWI (e.g. the reconstructed wavefield (a.k.a. extended source) approach of van Leeuwen and Hermann, 2013; Huang et al., 2016; Wang et al., 2016; Wang et al., 2017a; 2017b).

Apart from removing random noise and filtering the gradient, no other data processing would usually be necessary in the FWI process. Two-way wave equation modelling employed within the RTM engine will reproduce surface ghosts and all multiples, so these arrivals can be left in the field data, as they help to build a usable residual when compared to the modelled data. However, if a simple single-scattering Born method is employed for the reflection FWI, then multiples and ghosts will need to be removed from the field data (as with data for use with ray-tomography) as they will not have then been modelled.

Consequently, the overall workflow could be summarized as:

1. Estimate a ghost-free wavelet (perhaps for each source).
2. Perform initial model update using a cycle-skip avoidance technique, maybe with conventional ray-tomography or refraction FWI using a travel-time-based residual or shaping filter residual.
3. Refine the shallow model using refraction least-squares FWI increasing the frequency bandwidth so as to obtain better resolution (to depth $\sim 1/3$ max offset).
4. Update the deeper section using reflection FWI (which may require de-multipled data).

In addition, conventional ray-based tomography and interpretational picking of key high velocity contrast layers may be employed between any of these various stages. Remember that FWI attempts to match real and modelled shot data: it does not set-out to produce flat migrated gathers. Hence, we may still need to adjust velocities found by FWI with an additional constraint in the ‘image domain’, such as is done by image domain ray tomography. For data with very problematic near surface issues or sparse sampling (such as OBN data), surface wave or guided wave inversion could also be employed to constrain the near surface model (e.g. Socco et al., 2010; Boiero et al., 2013).

Conclusions

Contemporary implementations of FWI provide a mechanism for obtaining high-resolution velocity models suitable for use in migrating data from complex regions. In addition, the high resolution models can also be employed as constraints for impedance inversion (Cobo et al., 2018; Jones et al., 2018).

However, in Tarantola’s original work (1984), he did not conceive of using FWI as a means merely of providing a better migration model. His vision was far more ambitious, namely to circumvent the need for separate migration and elastic impedance inversion completely, by extending FWI in order to solve for more elastic parameters and higher frequencies (e.g. Jones, 2012; Routh et al., 2017). Unfortunately, this goal is currently just slightly beyond the capabilities of current computer power, but it is the most likely the way forward for development in the field of imaging and inversion, especially given that FWI employs the full wavefield: including all the multiples.

Acknowledgements

I would like to thank Jonathan Shewchuck for permission to use his images, to Peter Rowbotham, John Brittan, Carlos Calderon, Jacques Leveille, and David Yingst, for helpful suggestions with the manuscript, and to ION for permission to publish this work.

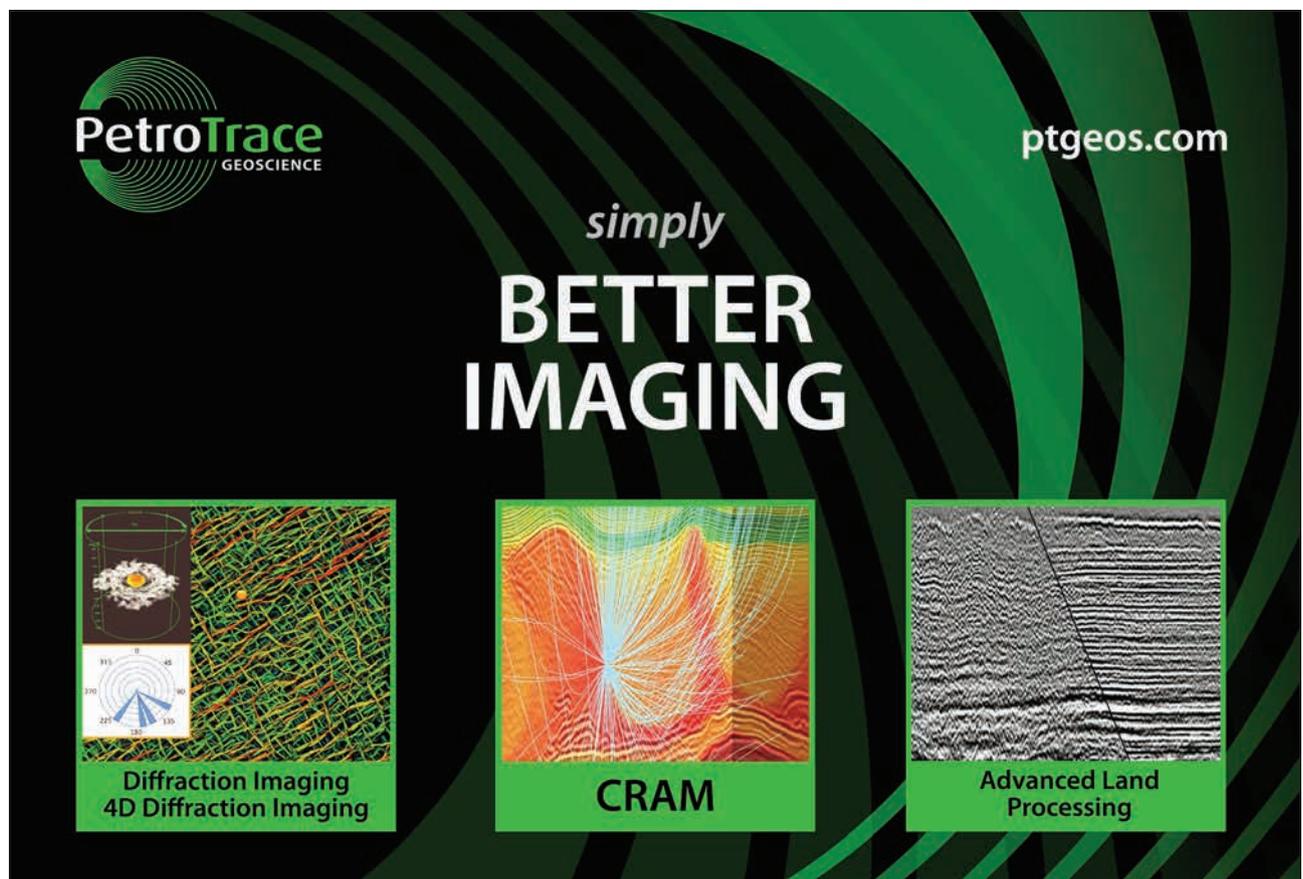
References

- Bancroft, J.C. [1997]. *A Practical Understanding of Pre/Poststack Vol.1*, SEG, USA.
- Bancroft, J.C. [2007]. *A Practical Understanding of Pre/Poststack Vol.2 (Pre stack)*, SEG, USA.
- Bevc, D., Fliedner, M.M., and Biondi, B., [2008]. Wavepath tomography for model-building and hazard detection. *78th SEG Annual Meeting*, SVIP2.7.
- Boiero, D., Strobba, C., Velasco, L. and Vermeer, P. [2013]. Guided waves- inversion and attenuation. *75th EAGE Conference & Exhibition*, Extended Abstracts.
- Claerbout, J.F. [1971]. Toward a unified theory of reflector mapping. *Geophysics*, **36** (3), 467-481.
- Claerbout, J.F. [1985]. *Imaging the Earth’s Interior*. Blackwell Scientific Publications, UK.
- Cobo, Y., Chi., S., Calderon-Macias, C. [2018] Quantitative interpretation from full-waveform inversion in a deep water GOM dataset. *80th EAGE Conference & Exhibition*, Extended Abstracts.
- Douma, H., Yingst, D., Vasconcelos, I. and Tromp, J. [2010]. On the connection between artifact filtering in reverse-time migration and adjoint tomography. *Geophysics*, **75**, S219–S223.
- Farmer, P., Jones, I.F., Zhou, H., Bloor, R. and Goodwin, M.C. [2006]. Application of Reverse-time Migration to Complex Imaging Problems. *First Break*, **24** (9), 65-73.
- Fu, L., Li, J., Dutta, G., Lu, K., Feng, S., Zhang, S., Luo, Y. and Schuster, G. [2017]. Wave Equation Inversion of Skeletonized Data. Accessed: 16/01/2019, http://csim.kaust.edu.sa/files/report/17_seg/Schuster.Workshop.2017.pptx
- Huang, H., Symes, W. and Nammour, R. [2016]. Matched source waveform inversion: Volume extension. *86th SEG Annual International Meeting*, Expanded Abstracts., 1364–1368.
- Jiao, K., Sun, D., Cheng, X. and Vigh, D. [2015]. Adjustive full waveform inversion. *85th SEG Annual International Meeting*, Expanded Abstracts, 1091–1095,
- Jones, I.F. [2018]. *Velocities, Imaging, and Waveform Inversion (The evolution of characterizing the Earth’s subsurface)*. EAGE, The Netherlands.
- Jones, I.F., Singh, J., Greenwood, S., Chigbo, J., Cox, P. and Hawke, C. [2018], High resolution impedance estimation using refraction and reflection FWI constraints: the Fortuna region, offshore Equatorial Guinea. *First Break*, **36** (11), 39-44.
- Jones, I.F., Kobylarski, M. and Brittan, J. [2017]. Tutorial: RTM Image Enhancement; bad models and mad physics. *First Break*, **35** (3), 31-38.
- Jones, I.F. [2014]. Tutorial: migration imaging conditions. *First Break*, **32** (12), 45-55.
- Jones, I.F. and Davison, I. [2014]. Seismic imaging in and around salt bodies. *SEG Interpretation*, **2**, SL1-SL20.
- Jones, I.F. [2012]. *From imaging to inversion*. SEG 2012 European honorary lecture, accessed: 12/01/2018, <https://seg.org/Education/SEG-on-Demand/id/4109/honorary-lecture-recordings-from-imaging-to-inversion>
- Jones, I.F. [2010]. *An introduction to velocity model building*. EAGE, The Netherlands.
- Lailly, P. [1983]. The seismic inverse problem as a sequence of before stack migrations: Conference on inverse scattering: Theory and application. *SIAM*, 206–220.

- Liu, F., Zhang, G., Morton, S.A. and Leveille, J.P. [2011]. An effective imaging condition for reverse-time migration using wavefield decomposition. *Geophysics*, **76**, S29–S39.
- Maharramov, M., Baumstein, A.I., Tang, Y., Routh, P.S., Lee, S. and Lazaratos, S.K. [2017]. Time-domain broadband phase-only full-waveform inversion with implicit shaping. *87th SEG Annual International Meeting*, Expanded Abstracts, 1297-1301.
- Mora, P. [1988]. Elastic wavefield inversion of reflection and transmission data. *Geophysics*, **53**, 750–759.
- Plessix, R.-E. [2007]. A Helmholtz iterative solver for 3-D seismic-imaging problems. *Geophysics*, **72**, SM185-194.
- Plessix, R.-E. and Perkins, C. [2009]. 3D full-waveform inversion with a frequency-domain iterative solver. *71st EAGE Conference & Exhibition*, Expanded Abstracts, U039.
- Plessix, R.-E., Baeten, G., de Maag, J., Klaassen, M., Zhang, R. and Tao, Z. [2010]. Application of acoustic full waveform inversion to a low-frequency large-offset land data set. Expanded Abstract, *80th SEG Annual International Meeting*, Expanded Abstracts.
- Pratt, R.G., Song, Z.-M., Williamson, P. and Warner, M. [1996]. Two-dimensional velocity models from wide-angle seismic data by wavefield inversion. *Geophysical Journal International*, **124**, 323-340.
- Pratt, R.G., Gao, F., Zelt, C. and Levander, A. [2002]. A comparison of ray-based and waveform tomography – implications for migration. *64th EAGE Conference & Exhibition*, Expanded Abstracts, B023.
- Ramos-Martinez, J., Crawley, S., Zou, Z., Valenciano, A., Klochikhina, E. and Chemingui, N. [2016]. A robust FWI gradient for high-resolution velocity model building. *87th SEG Annual International Meeting*, Expanded Abstracts, 1258-1262.
- Routh, P., Neelamani, R., Lu, R., Lazaratos, S., Braaksma, H., Hughes, S., Saltzer, R., Stewart, J., Naidu, K., Averill, H., Gottumukula, V., Homonko, P., Reilly, J. and Leslie, D. [2017]. Impact of high-resolution FWI in the Western Black Sea: Revealing overburden and reservoir complexity. *The Leading Edge*, **36**(1), 60–66.
- Schuster, G.T. [1993]. *Decomposition of the waveform gradient into short and long wavelength components*. Annual Report, Utah Tomography and Modelling/Migration Consortium, 79-88.
- Schuster, G.T. [2017]. *Seismic Inversion. Investigations in Geophysics #20*. SEG, USA.
- Shah, N., Warner, M., Nango, T., Umpleby, A., Stekl, I., Morgan, J. and Guasch, L. [2012]. Quality assured full-waveform inversion: Ensuring starting model adequacy. *82nd SEG Annual International Meeting*, Expanded Abstracts.
- Shen, X., Ahmed, I., Brenders, A., Dellinger, J., Etgen, J. and Michell, S. [2017]. Salt model building at Atlantis with Full Waveform Inversion. *87th SEG Annual International Meeting*, Expanded Abstracts.
- Shewchuck, J.R. [1994]. *An Introduction to the Conjugate Gradient Method Without the Agonizing Pain: Edition 1 ¼*. <https://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf>.
- Sirgue, L. and Pratt, R.G. [2002]. Feasibility of Full Waveform Inversion Applied to Sub-basalt Imaging. *64th EAGE Conference & Exhibition*, Expanded Abstracts, P177.
- Sirgue, L. and Pratt, R.G. [2004]. Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies. *Geophysics*, **69**, 231-248.
- Sirgue, L., Barkved, O.I., Van Gestel, J.P., Askim, O.J. and Kommedal, J.H. [2009]. 3D Waveform Inversion on Valhall Wide-azimuth OBC. *71st EAGE Conference & Exhibition*, Extended Abstracts, U038.
- Sirgue, L., Barkved, O.I., Dellinger, J., Etgen, J., Albertin, U. and Kommedal, J.H. [2010]. Full waveform inversion: the next leap forward in imaging at Valhall. *First Break*, **28**, 65–70.
- Sirgue, L., Denel, B. and Gao, F. [2011]. Integrating 3D Full Waveform Inversion into Depth Imaging Projects. *81st SEG Annual International Meeting*, Expanded Abstracts.
- Socco, L.V., Foti, S. and Boiero, D. [2010]. Surface-wave analysis for building near-surface velocity models - Established approaches and new perspectives. *Geophysics*, **75** (5), P. 75A83–75A102.
- Symes, W.W. [2008]. 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.
- Tarantola, A. [1987]. *Inverse problem theory*. Elsevier, The Netherlands.
- Tarantola, A. [1988]. Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation. *Pure and Applied Geophysics*, **128** (1–2), 365–399.
- van Leeuwen, T. and Hermann, F. [2013]. Mitigating local minima in full-waveform inversion by expanding the search space. *Geophysical Journal International*, **195**(1).
- Vigh, D. and Starr, E.W. [2008]. 3D prestack plane-wave, full-waveform inversion. *Geophysics*, **73** (5), VE135–VE144.
- Vigh, D., Starr, E.W. and Kapoor, J. [2009]. Developing earth models with full waveform inversion. *The Leading Edge*, **28**, 432–435.
- Vigh, D., Jiao, K., Cheng, X., Sun, D. and Lewis, W. [2016]. Earth-model-building from shallow to deep with full-waveform inversion. *The Leading Edge*, **35**, 1020–1030.
- Vigh, D., Jiao, K., Cheng, X., Sun, D. and Zhang, L.X. [2017]. Seeing below the diving wave penetration with full waveform inversion. *87th SEG Annual International Meeting*, Expanded Abstracts, 1466-1470.
- Virieux, J. and Operto, S. [2009]. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, **74**, WCC1-WCC26.
- Virieux, J., Asnaashari, A. Brossier, R. Métivier, L., Ribodetti, A. and Zhou, W. [2017]. An introduction to full waveform inversion. In: Grechka, V. and Wapenaar, K. (Eds.), *Encyclopedia of Exploration Geophysics*. SEG, R1-1-R1-40.
- Wang, C., H. Delome, C. Calderon, D. Yingst, J. Leveille, R. Bloor, and P. Farmer, [2011]. Practical strategies for waveform inversion. *81st SEG Annual International Meeting*, Expanded Abstracts, 2534-2538.
- Wang, C., Yingst, D., Farmer, P. and Leveille, J. [2016]. Full-waveform inversion with the reconstructed wavefield method. *86th SEG Annual International Meeting*, Expanded Abstracts, 1237-1241.
- Wang, C., Yingst, D., Farmer, P., Jones, I., Martin, G. and Leveille, J. [2017a]. Reconstructed full waveform inversion with the extended source. *87th SEG Annual International Meeting*, Expanded Abstracts, FW15.1.
- Wang, C., Yingst, D., Farmer, P., Jones, I., Martin, G. and Leveille, J. [2017b]. Reconstructed wavefield imaging with the extended source. *87th SEG Annual International Meeting*, Expanded Abstracts, SPMI E-P1.3.

- Wang, C., Farmer, P., Yingst, D., Jones, I., Calderon, C. and Brittan, J. [2018]. Traveltime based reflection full waveform inversion. *80th EAGE Conference & Exhibition*, Extended Abstracts.
- Warner, M., Umpleby, A., Stekl, I. and Morgan, J. [2010]. 3D full-wavefield tomography: imaging beneath heterogeneous overburden. *72nd EAGE Conference & Exhibition*, Extended Abstracts.
- Warner, M., Nangoo, T., Shah, N., Umpleby, A., Morgan, J. [2013]. Full-waveform inversion of cycle-skipped seismic data by frequency down-shifting. *83rd SEG Annual International Meeting*, Expanded Abstracts, 903-907.
- Warner, M.R. and Guasch, L. [2014a]. Adaptive waveform inversion: Theory. *76th EAGE Conference & Exhibition*, Extended Abstracts.
- Warner, M.R. and Guasch, L. [2014b]. Adaptive waveform inversion. *84th SEG Annual International Meeting*, Expanded Abstracts, 1089–1093.
- Warner, M.R. and Guasch, L. [2015]. Robust adaptive waveform inversion. *85th SEG Annual International Meeting*, Expanded Abstracts, 1059–1063.
- Warner, M.R. and Guasch, L. [2016]. Adaptive waveform inversion: Theory. *Geophysics*, **81** (6), R429- R445.
- Worthington, M.H. [1984]. An introduction to geophysical tomography. *First Break*, **2**, 20–26.
- Yang, Y. and Engquist, B. [2018]. Analysis of optimal transport and related misfit functions in full-waveform inversion. *Geophysics*, **83** (1), A7-A12.
- Yingst, D., Wang, C., Park, J., Bloor, R., Leveille, J.P. and Farmer, P. [2011]. Application of time domain and single-frequency waveform inversion to real data. *73rd EAGE Conference & Exhibition*, Extended Abstracts.
- Zhang, J. [1999]. Fat ray tomography with optimal relaxation factor. *69th SEG Annual International Meeting*, Expanded Abstracts.

ADVERTISEMENT



PetroTrace
GEOSCIENCE

ptgeos.com

simply
**BETTER
IMAGING**

**Diffraction Imaging
4D Diffraction Imaging**

CRAM

**Advanced Land
Processing**