By Andrew Long (Andrew.Long@pgs.com)

Full Waveform Inversion (FWI)

I attempt to explain FWI in accessible terms by presenting the critical steps for the methodology to yield useful results, and refer to some common ambitions of the many industry and academic implementations: 1. Setting up a scheme that adjusts the differences between modeled and measured shot gathers so that the subsequent steps in the iterative workflow are stable, 2. Isolating only the seismic information that is relevant to the subsurface parameters being isolated, and 3. Guiding the iterative recovery of an accurate subsurface model in a manner that reduces spurious artifacts.

If you're a strong mathematical geophysicist this is probably below your competency, but if you can tolerate some basic algebra and have an interest in what FWI is about, then hopefully this might help you complete the initial journey. This article is written to accompany the three-part TechByte webinar series on FWI.

What is FWI?

Full Waveform Inversion (FWI) estimates a spatially-variable model of subsurface parameters; most commonly a model of the interval velocity, using an iterative data fitting workflow. Using an initial (velocity) model estimate, simulated data (in the form of synthetic shot gathers) are generated using full wavefield modeling in a manner that attempts to incorporate the most relevant propagation effects for the subsurface location.

FWI forward modeling typically uses an acoustic approximation to the wave equation. Elastic FWI versions are in the early stages of development, but are not considered here.

As discussed below, key steps include, but are not limited to the following:

- Setting up an 'objective function' that describes how well a model estimate represents the seismic measurement. 'Optimization' of the objective function is based on a 'minimal variation principle'; which means that iterative changes to the model parameters will cause the modeled data to converge to the measured data.
- Modifying the methodology used for two way wave equation-based seismic migration to recover updates to the model parameters without contamination from the reflectivity information sought in traditional seismic migration.
- 'Regularization' of the objective function in an attempt to smooth noisy outliers in the model that prevent convergence, whilst preserving legitimate (geological) edges.

These considerations drive much of the R&D being published at technical conferences as any failure to properly implement these steps will cause FWI to 'fail spectacularly' (estimate an often wildly inaccurate model).

I begin by illustrating the similarities between Reverse Time Migration (RTM) and FWI, as both depend upon back-propagation of seismic wavefields to yield 'images' for shot gathers. For RTM, the images are of seismic amplitudes, and for FWI the images are used to compute *updates* to model parameters such as interval velocity. Note that FWI pursues several iterations of this pursuit until the updates no longer contribute meaningful information to the model. I then describe the basic mechanics of the 'gradient-based inversion' used within FWI, before elaborating on a few of the highest-profile industry methods used to make FWI robust.

FWI is Similar to RTM

FWI is similar to Reverse Time Migration (RTM), in that forward-propagated seismic wavefields from discrete source locations are correlated with back-propagated seismic wavefields from the associated receiver locations (the so-called 'adjoint sources'), and a form of seismic image is computed.

As shown in **Figure 1**, the RTM algorithm creates the source wavefields (upper row) by modeling seismic waves propagating forward in time, for each known source location, with a simple source wavelet, and a gridded subsurface velocity model with vertical units of depth. In the lower row, the recorded shot gather is injected into the model at all receiver locations (the upper part of the front panel), beginning with the largest time sample, and running backwards in time ('back-propagated'). 'Wavefield snapshots' are stored in memory for each sample of the respective forward-propagated source wavefield and back-propagated receiver wavefield, and an imaging condition extracts an image from appropriate pairs of source and receiver wavefields wherever the two wavefields are kinematically equivalent at the various subsurface locations—the timings, but not necessarily the amplitudes. In other words, the wavefield originating from the source is incident upon each reflector at the same location and instant that the reflected receiver wavefield reflects from that reflector. Note this traditional approach to migration assumes there are no multiples in the data, no diving waves and no refractions, whereas the modeling in FWI can include these wavefields.

The particular value of the time, t, is immaterial to the existence of the reflection point at each spatial location, only that the wavefields coincided at some time, t. To determine a time-coincident amplitude in both the source and receiver wavefields, the algorithm cross-correlates the two wavefields as a function of time for all spatial subsurface locations. The non-zero value of the cross-correlation coefficient is stored at the subsurface location, this process is repeated for all source and receiver wavefield snapshots, and the results are summed to produce the 'partial' image for each shot gather. Finally, all partial images from each shot gather are summed to yield the final migrated (stack) volume.

It is explained below that the 'image' produced for each shot in FWI has an entirely different physical meaning to the seismic amplitudes recovered during RTM.

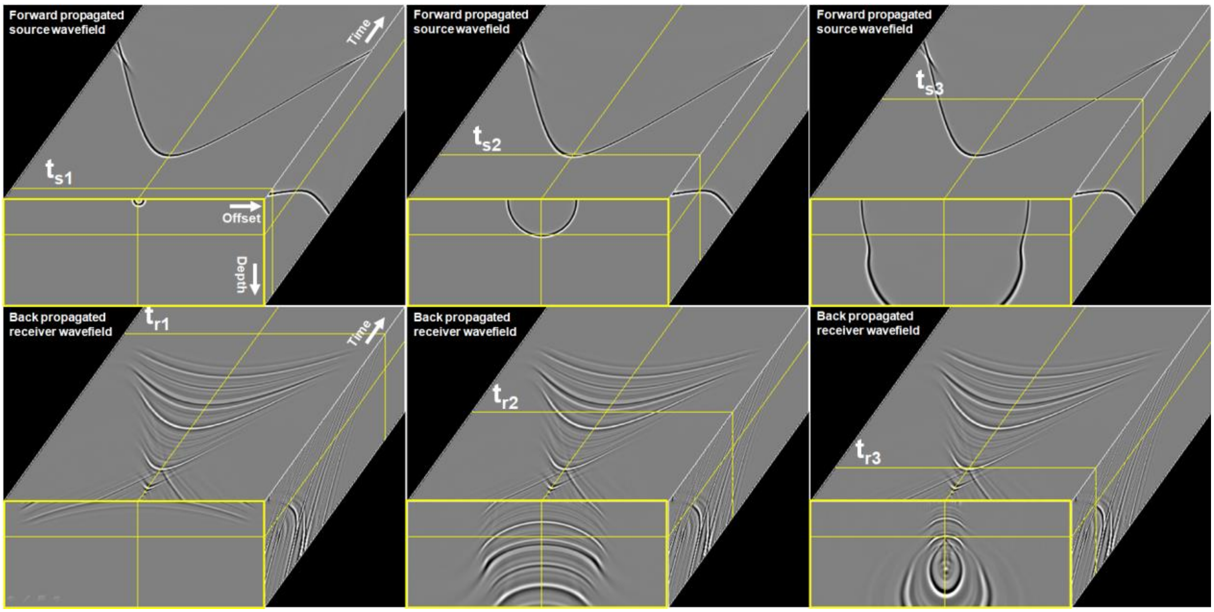

Figure 1. Wavefield propagation snapshots from RTM at arbitrary moments in time. For both rows of figures, the front panel within the yellow polygon is offset vs. depth, the top panel is offset vs. time, and the side panel is depth vs. time. (upper row) Forward-propagation of a source wavefield where a split-spread shot gather is simulated. The horizontal yellow line on the top panel indicates the sample time of the synthetic shot gather corresponding to the source wavefield in the front panel. (lower row) Back-propagation of the recorded shot gather. The horizontal yellow line on the top panel indicates the sample time of the recorded shot gather that is being injected into the velocity model at all the receiver locations. Courtesy of Paul Sava, Colorado School of Mines.

Inversion, Objective Functions and the Basic FWI Workflow

Seismic Inversion: Objective Functions

A common practical task in seismic processing and imaging is to take observed data, presented as a vector matrix of numbers, $\mathbf{d}_{obs}$, and fit that measured data to some modeled data, $\mathbf{d}_{mod}$, by the adjustment of components in the model parameters, $\mathbf{m}$.

Using common algebra, we seek $\mathbf{d}_{\mathbf{obs}} \approx \mathbf{d}_{\mathbf{mod}} = \mathbf{Lm}$, where $\mathbf{L}$ is an operator that transforms the model, $\mathbf{m}$, into the modeled data, $\mathbf{d}_{\mathbf{mod}}$. More commonly, it is said that $\mathbf{L}$ maps the seismic data from the model space to the data space.

It is typically the case in seismic processing when using linear operators that the number of observed data values is less than the number of (desired) model values (an 'undetermined problem'), so $\mathbf{L}$ is a wide rectangular matrix, and cannot be solved directly. However, an iterative method ('inversion') can be used as follows. Note that most seismic inversion problems are nonlinear, the problem is often linearized, and the final nonlinear solution is obtained through the iterative application of linearized solvers.

Inversion problems begin with the minimization of an 'objective function'; sometimes also called a 'cost function' or 'misfit function'. In the case of FWI, the objective function can be simplistically written as follows (I avoided the full details of the classical notation):

$$J(\boldsymbol{m}) = \sum_i \left\| d_{mod,i}(\boldsymbol{m}) - d_{obs,i} \right\| \tag{1}$$

where the vertical pairs of lines represent the L_2 norm of a vector. The physical meaning of a vector 'norm' is a measure of distance. The most commonly used norm for minimizing the FWI objective function is the L_2 norm, which is the square root of the sum of the squares of its components (the 'error', or difference between the measured and modeled seismic traces, computed on a sample-by-sample basis). Ideally, the vector norm will converge steadily to zero, or at least some small value that indicates the 'global minima' has been found (refer to **Figure 2**).

Any modification to the operator $\mathbf{L}$ (or its approximation) is called regularization, which makes the inverse mapping from the data space to the model space happen in a stable and, hopefully, unique manner. Regularization attempts to suppress singularities that make the problems ill-posed, which creates computational instabilities.

In practice, the measured data is incomplete and imperfect, so the objective function involves two data-fitting considerations, where ϵ balances the contribution of the two criteria in the objective function:

$$J(\boldsymbol{m}) = \sum_i \left\| d_{mod,i}(\boldsymbol{m}) - d_{obs,i} \right\| + \epsilon \left\| \boldsymbol{m} - \boldsymbol{m}_{ref} \right\| \tag{2}$$

Regularization is therefore defined by various model constraints, and is applied additively to the data fitting—hopefully without over-fitting the model (see also further discussion below).

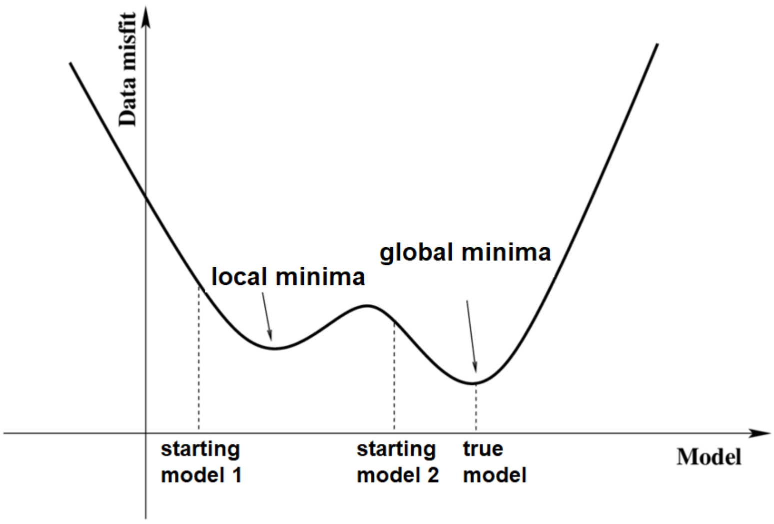

Figure 2. Schematic illustration of local vs. global minima.

Basic FWI Workflow

To translate FWI in the context of **Figure 3**, we minimize the difference between the modeled and observed shot gathers in an iterative manner by continually perturbing the model parameters, $\mathbf{m}$, with the ambition that this difference will converge as closely as possible to zero. Initially, the modeled shot gather will only contain a subset of the observed seismic events, and with different arrival times.

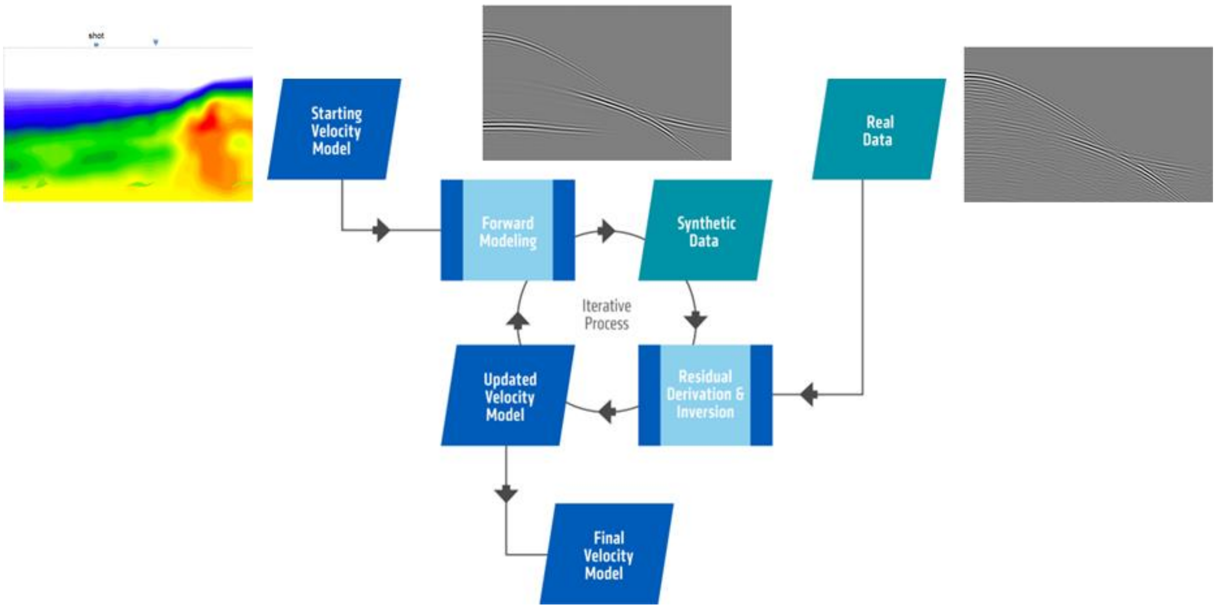

Figure 3. Schematic FWI workflow that includes an initial modeled shot gather from the smooth starting velocity model. An iterative evaluation of the objective function after each model update will determine whether the inversion has converged to a minima. Refer also to Figure 2.

In the first iteration of the FWI workflow the right-hand term in equation (2) will be zero, and the output from the objective function (the ‘residual’) is simply the difference between the initial modeled shot gather and the observed shot gather. The residual correspondingly represents that part of the measured data that is not taken into account by the current model.

In the manner of **Figure 1** and the implementation of RTM, the residual is back-propagated into the model, and the wavefield snapshots for the forward-propagated modeled source wavefield and the back-propagated residual wavefield are cross-correlated and summed to produce an image that represents perturbation wavefields scattered by the missing heterogeneities in the starting model. This image represents the mathematical gradient, and FWI pursued in this manner is referred to as a ‘gradient-based inversion’.

The concept of the gradient can be understood by thinking of the desired model vector, $\mathbf{m}$, as a topographical solution space (**Figure 2**) wherein the ‘global minima’ represents the true model. The gradient is the derivative of the objective function with respect to the model parameter, and the gradient vector can be interpreted as the ‘direction and rate of fastest increase’ (or fastest decrease in the case of the negative gradient).

In physical terms, an appropriately scaled version of the FWI gradient derived by application of an imaging condition to the wavefield snapshots comprises the update to the model for that iteration, and the update is added to the model. This iterative workflow continues until the objective function in (2) has robustly (in theory) converged towards the global minimum, the gradient has effectively vanished, and consequently it is assumed that something close to the true model has been recovered.

Three challenges to the convergence of the objective function in (2) are now briefly considered in the context of how they may cause FWI to either simply not converge at all, or converge to local minima with resultant significant errors in the recovered model:

1. Cycle skipping: Misalignment between modeled and measured shot gathers that is larger than half of the dominant period (wavelength)
2. Inappropriate imaging conditions: Contamination of the model updates with reflectivity information
3. Inadequate regularization of the objective function: Noisy model updates, failure to recover legitimate sharp boundaries, or leakage between model parameters

I also note that coherent noise in the observed data may contaminate the objective function, and it is obviously not desirable to minimize any such contribution with erroneous model updates.

Mitigation of Cycle Skipping

There is a critical challenge to FWI known as 'cycle skipping' that must be accounted for within each FWI iteration before the model update can be computed. If the time misalignment between the modeled and observed data is more than half a cycle (wavelength) for any considered frequency, the objective function can easily converge to a local minima, and the iterative process will terminate prematurely. As shown in the simple example of **Figure 4**, when the incorrect residual is back-propagated into the model, the computed update may be wildly noisy. Equally problematic is the fact that the errors in the estimated model may appear to be 'geological' and 'high frequency' in nature.

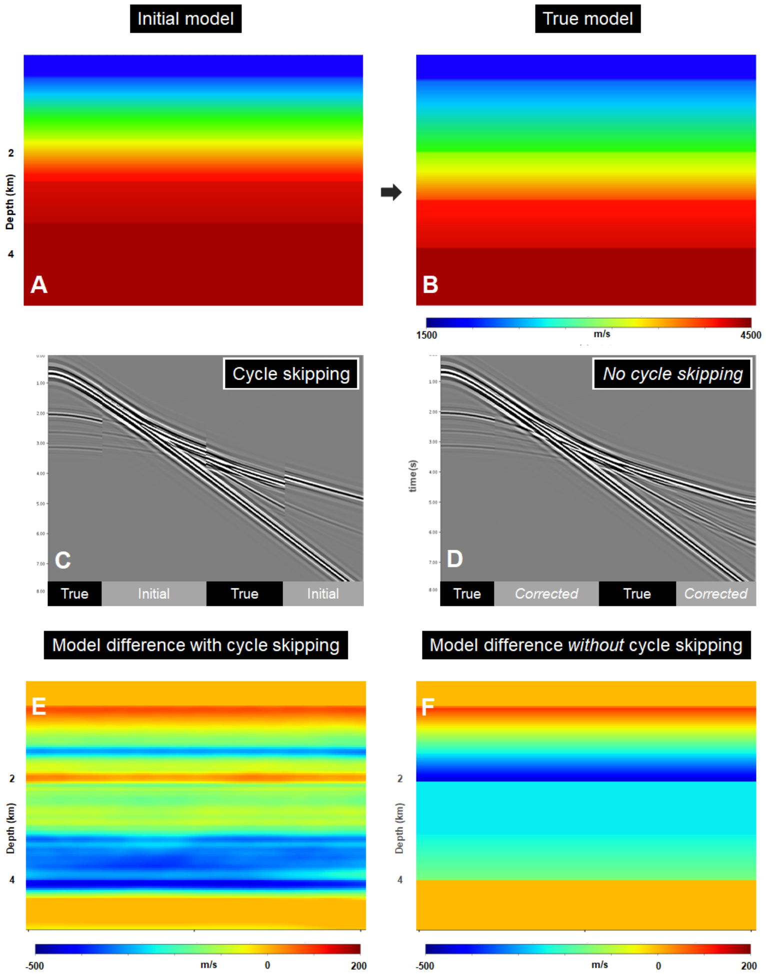

Figure 4. (A) Initial velocity model; (B) True velocity model; (C) Modeled shot gather with alternating offset ranges to show the time shifts associated with the two velocity models; (D) Alternating offset ranges with corrected time shifts; (E) Velocity model updates recovered if cycle skipping us not accounted for; and (F) Correct velocity model updates (the difference between the true and initial model).

Early FWI implementations pursued a smoother objective function by using a starting model that was built with high effort to be as close as possible to the final model (e.g. reflection tomography). As cycle skipping becomes more problematic with higher frequencies, FWI usually starts at the lowest frequency in the data that has coherent phase, and then progressively incorporating higher frequencies into the modeled data.

Recent industry and academia FWI efforts to account for cycle skipping go by many names, but each attempts to evaluate the misfit between simulated and measured data in a more robust manner, and then modify the objective function to reduce the kinematic errors, penalize higher amplitudes in the residual associated with the kinematic errors, and so. In this context, the following labels used to describe FWI implementations have common ambitions: "Time warping, time-lag, data-adaptive convolutional matching filters, dynamic warping, dynamic matching, weighted objective function, adjustive, well-constrained, Wasserstein distance / W2 norm", and so on. Note that the measured data is adjusted to reduce the misfit, rather than the modeled data, as a change to the modeled data would imply a change in model parameters that is not valid.

Removing the problem of cycle skipping enables both a simplified starting model for FWI, and higher frequencies.

Including Reflections into FWI

The historically most common FWI implementations rely upon diving wave and refraction information; which only provide model updates to the deepest turning point: about one-sixth to one-third of the maximum available offset. Traditional towed streamer lengths of 8-10km therefore cannot provide model updates for depths larger than about 3km (or less). This has been an important driver in the use of Ocean Bottom Node (OBN) surveys in salt-prone areas where deep model updates are desired to reduce the substantial drilling risks (see discussion below).

Reflection FWI will recover much deeper model updates for a given offset range than diving wave FWI if successful, however, most industry implementations of FWI will only attempt to subsequently incorporate reflections if the diving wave FWI can yield an appropriately high-quality background model capable of generating the synthetic reflections. Part of the traditional reliance upon diving waves only is also because the event phase is the most critical parameter when attempting to avoid cycle skipping and noisy model updates during diving wave FWI. Some industry implementations consequently penalize the impact of event amplitudes within the objective function for diving wave FWI.

As illustrated in **Figure 5**, the use of a traditional imaging condition when computing the reflection FWI gradient will include a migration isochrone that contributes high wavenumber (i.e. reflectivity) information to the model updates. **Figure 5** shows 'velocity sensitivity kernels', which represents the sensitivity of the objective function to a change in one model parameter whilst holding all other model parameters constant. In other words, it shows where model updates can be expected for a given offset and reflector depth. As illustrated in the right side of **Figure 5**, the gradient computation of Ramos-Martinez et al. (2016) allows the migration isochrone contribution to be isolated and removed, and assuming that cycle skipping is managed, residuals that match the modeled data with an error less than half a period will be back-projected constructively over the first Fresnel zone, updating the large wavelengths (i.e. low wavenumbers) of the model. Furthermore, velocity updates are no longer forced to coincide with impedance layer contrasts.

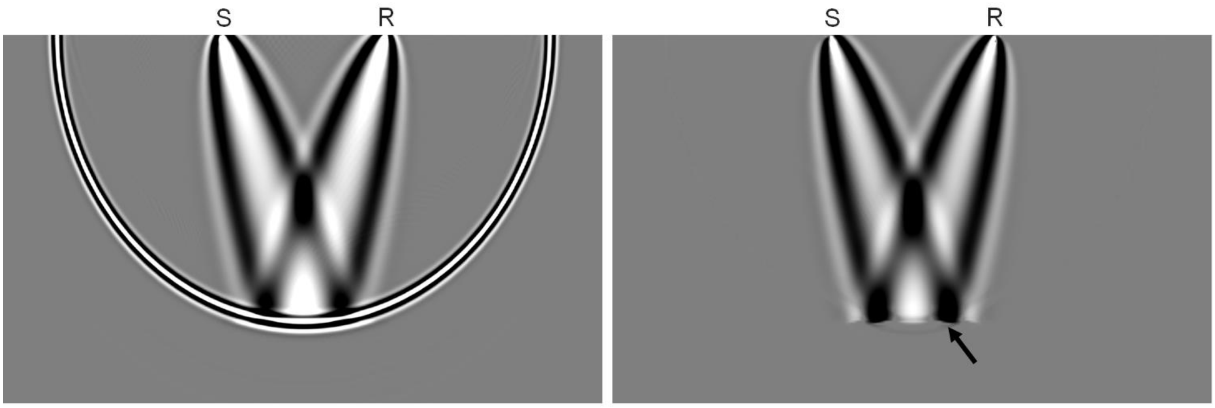

Figure 5. Sensitivity kernels for reflection FWI. The reflector depth is 4km, the source-receiver offset is 2.5km, and the background velocity is 4 km/s. (left) Conventional imaging condition that includes the unwanted migration isochrones; and (right) PGS imaging condition that isolates the low wavenumber contributions to the gradient.

The remaining challenge to reflection FWI is the difficulty of generating reflections from a smooth initial velocity model without a knowledge of the density model, hence the traditional sequence of diving wave FWI followed by reflection FWI. The 'vector reflectivity' solution shown in **Figure 6** uses a new parameterization of the variable density wave-equation and an *a priori* estimate of reflectivity to fast-track the modeling and gradient computation in reflection FWI. As the reflectivity is derived from the seismic data, the reflection FWI no longer needs a speculative density function. Furthermore, in addition to reflections, the new method can synthesize refractions and turning waves, which is beyond the capability of the currently available industry methods.

Overall, steady progress is being made towards reflection FWI being a robust solution for recovering deep high-resolution velocity updates in all geological settings from offsets typical of towed streamer acquisition.

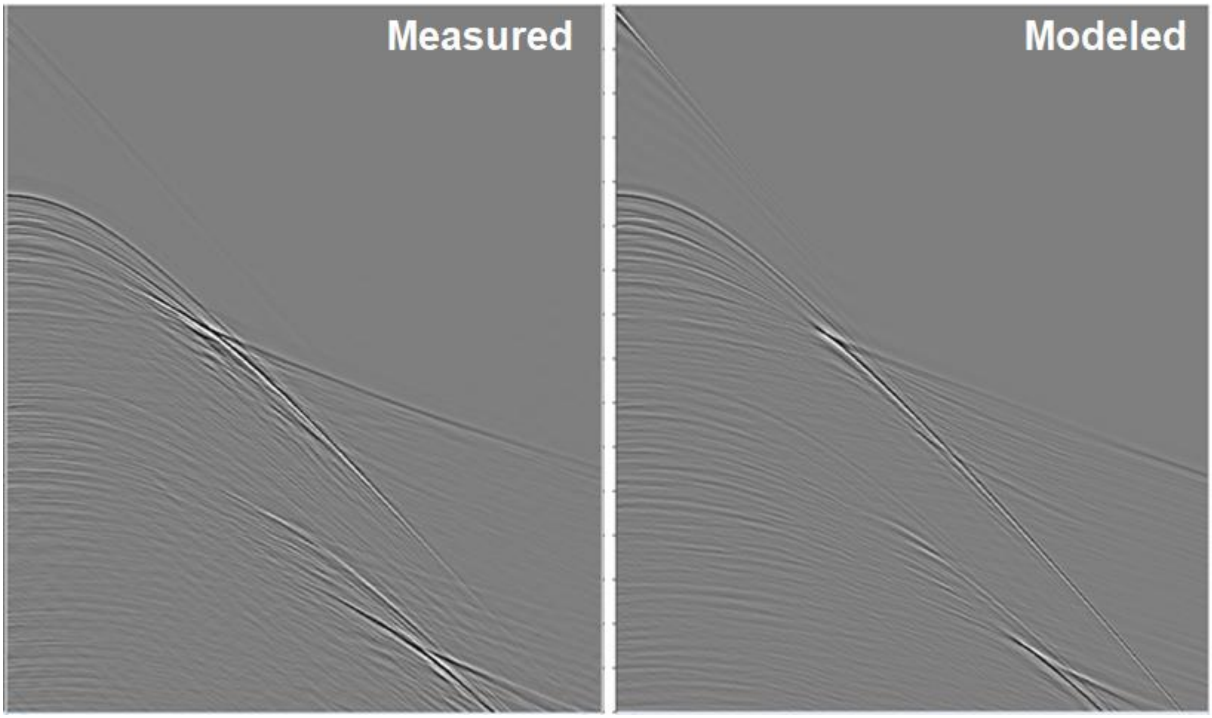

Figure 6. (left) Deghosted multisensor shot gather; and (right) Modeled shot gather using the 'vector reflectivity' method of Whitmore et al. (2020).

Regularization

With reference to equation (2) introduced earlier, regularization attempts to smooth noisy outliers in the model that prevent convergence, whilst preserving legitimate (geological) edges. Stated another way, regularization seeks to improve the generalizability of the model by minimizing the residual over all samples, i.e. avoids over-fitting the model. In practice, this means a penalty is imposed on the complexity of the model, and may, for example, involve imposing prior distributions on model parameters, seeking sparseness in the model (i.e. using only the largest coefficients), reducing the total variation in the model (i.e. introducing a non-linear local smoothing), and so on.

In practice, regularization can enable reflection FWI to robustly recover deep model updates in environments with high contrast features; whether fast bodies such as salt, volcanics or carbonates, or slow bodies such as oozes or gaseous sediments. Regularization can also reduce leakage between parameters in multi-parameter FWI.

Figure 7 is a simple synthetic example that uses the BP salt model to illustrate how appropriate regularization terms in the objective function can improve the recovery of deep and high contrast features in the model.

Other FWI Considerations

All discussion here assumed that the acquisition geometry provided the necessary platform for the various implementations of FWI to be robust. The highest-profile industry discussion today related to survey design involves the use of OBN surveys as a complement to towed streamer 3D surveys, i.e. 'hybrid streamer-OBN' acquisition. Rather than use an OBN density that enables high-quality seismic imaging, a reduced cost 'sparse' OBN distribution is used with the explicit ambition of building a higher quality FWI velocity model than is possible using conventional single-vessel streamer lengths in the range of 8-10km. An implicit driving assumption also seems to be that reflection FWI implementation is still immature within the industry, so ultra-long offsets are necessary for diving wave FWI to recover deep model updates with high resolution. In addition to the fact that the sources and receivers are physically decoupled in an OBN survey, thereby enabling unlimited offset ranges (with appropriate source vessel planning), OBN receivers (at least the hydrophone, even if variable seafloor coupling creates unreliable particle motion data) have higher Signal-to-Noise Ratio (SNR) at ultra-low frequencies than towed streamer data. The SNR difference has not yet been universally documented in different seafloor settings, but higher SNR at ultra-low frequencies can reduce cycle-skipping effects when pursuing very deep model updates in sub-salt areas where drilling risk is particularly high. It can be noted that migration image quality is mostly driven by the low-wavenumber background velocity model. For salt provinces, the high-wavenumber migration term (i.e. the migration isochrones in the left side of **Figure 5**) and the sharpness of salt boundaries in an FWI velocity model have little impact upon kinematics during migration above about 8 Hz, which may be another factor in the growing use of OBN for FWI. I will address emerging OBN acquisition models in another Industry Insights newsletter.

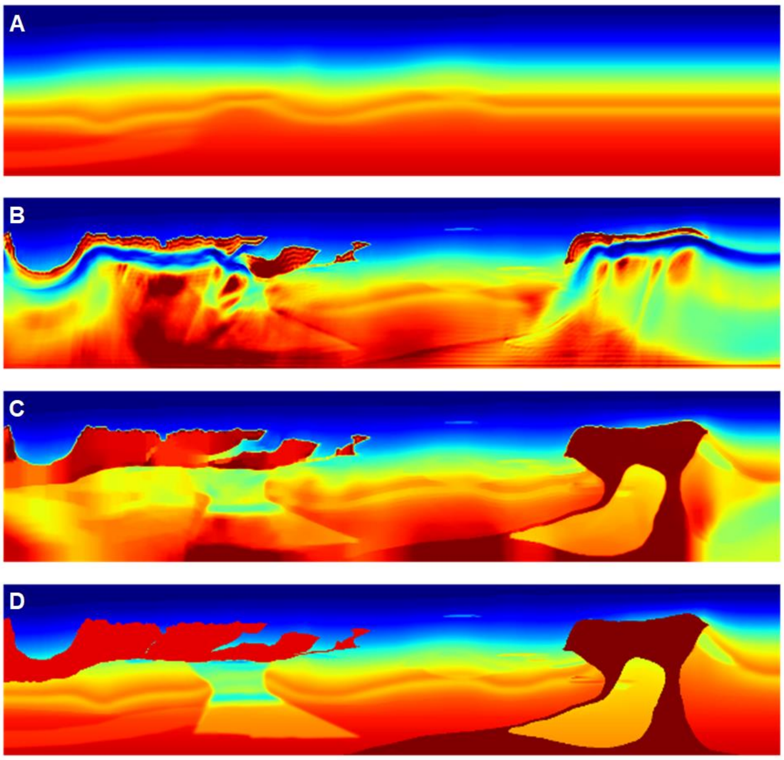

Figure 7. (A) Initial smooth velocity model; (B) Velocity model recovered without any regularization of the objective function; (C) Velocity model recovered with appropriate regularization of the objective function; and (D) Correct velocity model.

[The third part of the TechByte webinar series on FWI] briefly considers two high-profile aspects of how FWI may be extended to applications beyond velocity model building for more accurate seismic imaging:

- Using rock physics-based conditioning of FWI velocity model to build low frequency elastic impedance models for quantitatively accurate seismic reflectivity inversion, and
- Running the synthetic modeling in FWI to high frequencies approaching those present in migrated seismic reflectivity images.

I will address both considerations in future Industry Insights newsletters. The links between FWI and Least-Squares Migration (LSM) are particularly relevant to the pursuit of higher frequency model information, and data domain LSM similarly relies upon back-propagation of residuals. Furthermore, the back-projection of (non-linear) inversion errors is a key component of deep neural network training for machine learning applications; and clear mathematical links can be demonstrated to the methods discussed here.

There was no space here to consider multi-parameter FWI wherein other subsurface parameters such as anisotropy (most commonly the epsilon parameter), absorption (defined by the quality factor, Q), and any other parameter that can be robustly represented during the forward-propagation of the modeled shot gathers. In some geological settings, multi-parameter FWI may reduce non-uniqueness in the model related to coupling between the vertical velocity field and the anisotropy field, may compensate for certain phase dispersion effects, and so on, with the common ambition of improving near-surface geological detail in the model. It should be noted that velocity errors dominate the data misfit in the objective function, and other parameters introduce second-order errors.

Elastic FWI development is immature, but if successful, models for both the compressional and shear-wave velocity (Vp and Vs) may allow elastic RTM to better resolve certain lithologies; especially in areas affected by gas clouds. The direct recovery of elastic earth models from shot gathers remains something of a holy grail for the industry, and this will also be touched upon in the Industry Insights newsletter on LSM.

Time-lapse (4D) FWI can be used to monitor fluid migration during reservoir production with an improved spatial representation, and courtesy of changes in model parameters rather than changes in recorded data amplitudes. When appropriate rock physics-based calibrations are available, such an approach may also reduce the influence of survey geometry errors on the reservoir imaging.

Summary

FWI is typically executed as a gradient-based inversion scheme that has practical links to aspects of RTM migration of shot gathers. I discussed the goals common to most industry and academia implementations of FWI, irrespective of the marketing terminology used:

- Minimize an objective function that describes the difference between measured and modeled shot gathers, cycling through an iterative workflow that simultaneously attempts to update a model of subsurface parameters in a manner that is as close as possible to the truth.
- Regularization of the objective function seeks optimize the generalizability of the model(s) whilst preserving legitimate sharp boundaries in the model.
- The computation of the gradient in each inversion step involves the computation of a form of seismic image in a manner that should not force model updates and features to coincide with the impedance boundaries that generate reflections during seismic migration.

Suggested Reading Material

- Full Waveform Inversion: https://www.pgs.com/imaging/tools-and-techniques/full-waveform-inversion/
- PGS FWI Developments | A Seismic Leap to the Next Level: https://www.pgs.com/publications/feature-stories/pgs-fwi-developments--a-seismic-leap-to-the-next-level/
- Ramos-Martinez, R., Valenciano, A., Jiang, X., and Chemingui, N., 2019, Robust FWI updates in the presence of cycle skipping. 81st Conference & Exhibition, EAGE, Extended Abstracts, Tu_P01_08: https://www.pgs.com/globalassets/technical-library/tech-lib-pdfs/eage2019_ramos-marinez_etal_fwi.pdf
- Whitmore, N.D., Ramos-Martinez, J., Yang, Y., and, Valenciano, A., 2020, Full wavefield modeling with vector reflectivity: Conference & Exhibition, EAGE, Expanded Abstracts, We_Dome1_07.