

# Frugal full-waveform inversion: From theory to a practical algorithm

FELIX J. HERRMANN, IAN HANLON, RAJIV KUMAR, TRISTAN VAN LEEUWEN, XIANG LI, BRENDAN SMITHYMAN, and HANEET WASON, *University of British Columbia*  
 ANDREW J. CALVERT and MOSTAFA JAVANMEHRI, *Simon Fraser University*  
 ERIC TAKAM TAKOUGANG, *University of Western Australia*

As conventional oil and gas fields are maturing, our profession is challenged to come up with the next-generation of more and more sophisticated exploration tools. In exploration seismology this trend has led to the emergence of wave-equation-based inversion technologies such as reverse time migration and full-waveform inversion. While significant progress has been made in wave-equation-based inversion, major challenges remain in the development of robust and computationally feasible workflows that give reliable results in geophysically challenging areas that may include ultralow shear-velocity zones or high-velocity salt. Moreover, subsalt production carries risks that need mitigation, which raises the bar from creating subsalt images to inverting for subsalt overpressure.

Among the many challenges that wave-equation-based inversion faces, we focus on reducing the excessive computational costs of full-waveform inversion (FWI). We accomplish these cost reductions by using modern techniques from machine learning and compressive sensing. Contrary to many implementations of wave-equation-based inversion, we propose a methodology where we do not insist on using all data (i.e., looping over all sources) to calculate the velocity model updates. Instead, we rely on a formulation that calls for more data and more accuracy in the wave simulations only as needed by the inversion. This dynamic selection of data and accuracy leads to major savings, in particular in the beginning when we are far from reaching the solution. Because this approach reduces the computational costs significantly, we open the way to test different scenarios for the purpose of quality control or to include more sophisticated regularization. Without this cost reduction, both would be computationally infeasible.

The article is organized as follows. First, we introduce the basics of full-waveform inversion, followed by stochastic sampling techniques to reduce the computational costs. Next, we propose an adaptive scheme that dynamically selects the required sample size—i.e., the number of source experiments that partake in the inversion, and forward modeling accuracy. Recent results on the Chevron Gulf of Mexico data set are discussed next, followed by a discussion on challenges of FWI and possible roads ahead.

## Wave-equation-based inversion

Full-waveform inversion (FWI) is a parameter estimation problem, seeking Earth models that explain observed data typically collected in shot records. FWI is generally cast as an optimization problem where Earth models are found by minimizing the misfit between observed and simulated data—i.e., we solve the following optimization problem:

$$\min_{\mathbf{m}} \Phi(\mathbf{m}) = \sum_{i=1}^M \frac{1}{2} \|\mathbf{d}_i^{\text{obs}} - \mathbf{d}_i^{\text{syn}}(\mathbf{m})\|_2^2 \quad \text{where} \quad \mathbf{d}_i^{\text{syn}}(\mathbf{m}) = \mathbf{P}_i \mathbf{A}_i(\mathbf{m})^{-1} \mathbf{q}_i,$$

where

- $\Phi(\mathbf{m})$  is the objective function that needs to be minimized over the unknown model parameters collected in the vector  $\mathbf{m}$ ,
- $M$  is the number of (monochromatic) experiments,
- $\mathbf{m}$  is the model vector of length  $N$  with gridded medium parameters, e.g., the square of the gridded slowness,
- $i = 1 \cdots M$  is the experiment index,
- $\mathbf{d}_i^{\text{obs}}$  are the observed shot records,
- $\mathbf{d}_i^{\text{syn}}(\mathbf{m})$  are the synthetic shot records calculated from the current model  $\mathbf{m}$ ,
- $\mathbf{A}_i(\mathbf{m})$  is the discretized wave equation—e.g., the Helmholtz system,
- $\mathbf{q}_i$  represent the discretized source vectors,
- $\mathbf{P}_i$  is the detection operator extracting the synthetic shot record for the  $i^{\text{th}}$  source experiment at the receiver locations.

Solutions of the above minimization problem are typically found by using iterative optimization techniques, which use local derivative information to compute descent directions that minimize the objective  $\Phi(\mathbf{m})$ . At the bare minimum, these optimizations consist of the following steps for each experiment (Tarantola and Valette, 1982; Plessix, 2006): (1) data prediction by solving the wave-equation by inverting  $\mathbf{A}_i(\mathbf{m})$ , (2) calculation of the misfit between observed and predicted data, (3) solution of the time-reversed or adjoint wave equation using the data residual as the source term, (4) calculation of the model updates by correlating the forward and adjoint wavefields for each experiment, followed by a summation over the experiments yielding the gradient, and (5) computing the model update by properly scaling the gradient—e.g., by conducting a linesearch and by approximately inverting the Hessian. This leads to efficient algorithms because the steps before the summation can be carried out in parallel.

While the above procedure is largely responsible for current successes of full-waveform inversion to field data, numerous challenges remain. These include the following:

- dependence on accurate long-wavelength starting models and long-offset and low-frequency data to avoid getting trapped in local minima
- increasing size and depth of survey areas, which leads to exponential increases in the computational and storage costs
- reliance on accurate modeling of the wave physics, which becomes computationally extremely challenging when moving to visco/poro-elasticity

- lack of robust and versatile workflows for data sets from different geological areas

Numerous researchers have embarked on addressing these challenges. Because successful application of geophysical methods often depends on multiple runs to fine-tune a multitude of inversion parameters, our aim in this article is to speed up turnaround times and to automate the inversion as much as possible. We accomplish this by casting FWI as an optimization problem that is economic in its use of computational resources. We accomplish this by using recent developments in large-scale optimization in machine learning where the objective and gradients are calculated only up to a precision required by the algorithm to converge. In practice, this translates to reduced data sizes (number of terms in the sum of the above objective) and to reduced accuracy in wave simulations in the early stages of the inversion. In this way, we make FWI computationally manageable so resources can be spent on where it matters namely in figuring out parameter settings, (pre)-processing, and starting model sensitivities.

Essentially, our approach is based on the “intuitive” premise that data misfit and gradient calculations do not need to be accurate at the beginning of an iterative optimization procedure when the model explains the data poorly. We translate this idea to a concrete and practical algorithm where we control the errors by adaptively increasing the number of shots and the simulation accuracy. This intuition can perhaps be explained as follows. If we need to drive from A to B, it is generally not necessary to depart on one’s journey in the exact direction. Heading in the right direction first, following some general imprecise directions, is fine as long as the directions become more precise as we are reaching the target destination B.

**FWI by sample averaging and stochastic optimization**

Aside from lack of robustness with respect to starting models and its ability to capture the wave physics adequately, FWI

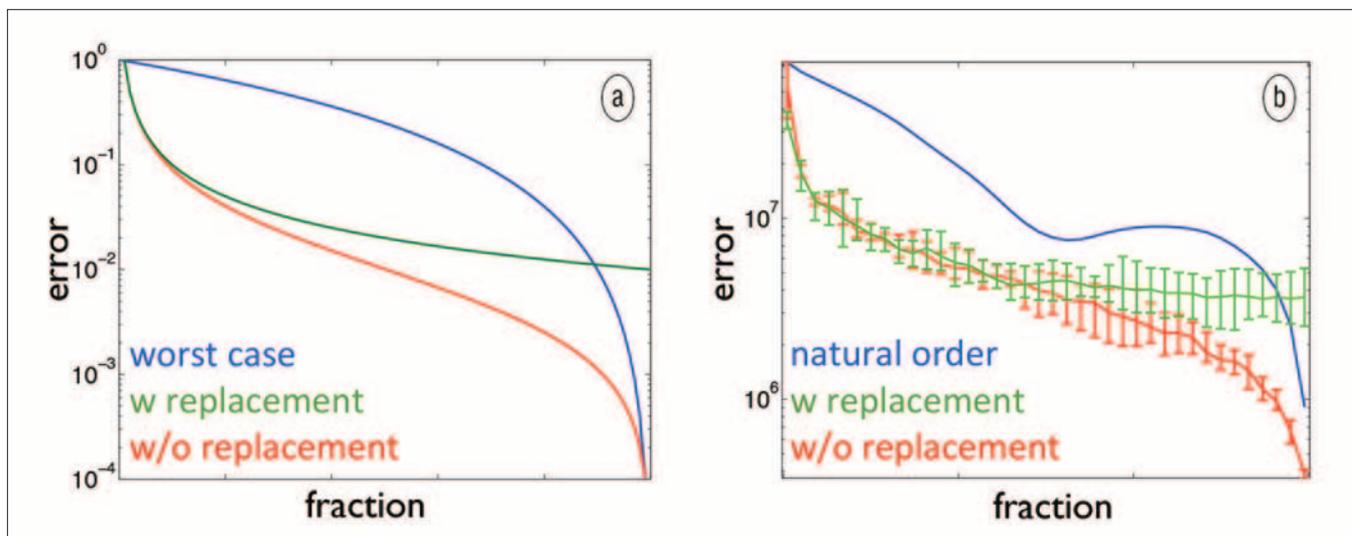
has been challenging computationally because the total work needed to calculate model updates scales linearly with the number of source experiments. To make matters worse, the number of source experiments and amount of work per experiment increase exponentially with the survey area. For that reason, we face with 3D FWI similar problems as researchers in supervised machine learning face, who are confronted with exponential growth in “big data.” Fortunately, researchers in both fields have been able to exploit the separable structure of their problems, namely the objective function and its gradient have the form of a sum over different experiments in both fields. Following recent work (Aravkin et al., 2012), this sum can be interpreted as an average over the experiments that can be approximated by its sample average—i.e., we have

$$\Phi(\mathbf{m}) \approx \frac{1}{K} \sum_{i \in J} \phi_i(\mathbf{m}; \mathbf{d}_i^{\text{obs}}, \mathbf{d}_i^{\text{syn}}) \quad \text{and} \quad \mathbf{g} = \nabla \Phi(\mathbf{m}) \approx \frac{1}{K} \sum_{i \in J} \nabla \phi_i(\mathbf{m}; \mathbf{d}_i^{\text{obs}}, \mathbf{d}_i^{\text{syn}})$$

where the sum now runs over a subset of experiments  $J$  with  $K = \text{size}\{J\} \ll M$  only. While this reduction in the sample size (i.e., the number of experiments) leads to a reduction in the computational costs, practical application of this approach calls for an answer to the following questions:

- What errors do we make when we work only with subsets of experiments, and how do these errors behave as a function of the size of these subsets?
- How should we select these subsets at each iteration of the optimization to ensure convergence and to minimize subsampling-related errors accumulated during the optimization?

Substantial theoretical progress has been made toward answering these questions. In the next sections, we summarize and translate these findings as they relate to the practicalities of FWI.



**Figure 1.** (a) Theoretical sampling errors as a function of the fraction of shot records with replacement (green), without replacement (red), and worst case (blue). (b) The same but now in practice for different random realizations of the subset selection. Because we do not know the worst case, we selected sources in a natural order.

Downloaded 09/27/13 to 134.7.248.132. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

**Randomized sampling without replacement**

The first step in the reduction of the computational costs of FWI is the selection of subsets of source experiments. Selection of these subsets leads to errors, and Figure 1 shows the behavior of the theoretical and empirical sampling errors of the gradient as a function of sample size divided by the total data size. In this plot, we compare the worst case scenario, where we sample the least informative experiments, with randomized sampling with and without replacement. As expected, the sampling errors for the randomized sampling decay as the fraction increases but this decay stalls if we allow the experiments to be selected more than once. This is the case when we select the source experiments with replacement. (Or, equivalently, when we generate new supershots by aggregating sequential into simultaneous shots with amplitude or phase encoding. See van Leeuwen and Herrmann, 2012, and references therein.)

However, if we select the experiments without replacement, the error continues to decay and the sampling eventually becomes deterministic as the fraction of selected sources goes to one. These theoretical predictions are confirmed by empirical findings also included in Figure 1. These findings also show that we can control the error of each gradients but the question “how does this subsampling error affect the convergence of FWI?” still needs to be answered. The field of stochastic optimization attempts to answer this question.

**Hybrid deterministic-stochastic optimization**

Even through randomly choosing small subsets of sources speeds up computations, it introduces unwanted random errors in the misfit and gradient calculations that may affect convergence of FWI and the quality of the end result. By choosing independent random subsets for each gradient calculation—an approach known as stochastic gradient descent widely used in data-intensive fields such as machine learning—convergence can be greatly improved because biases in the gradient are removed.

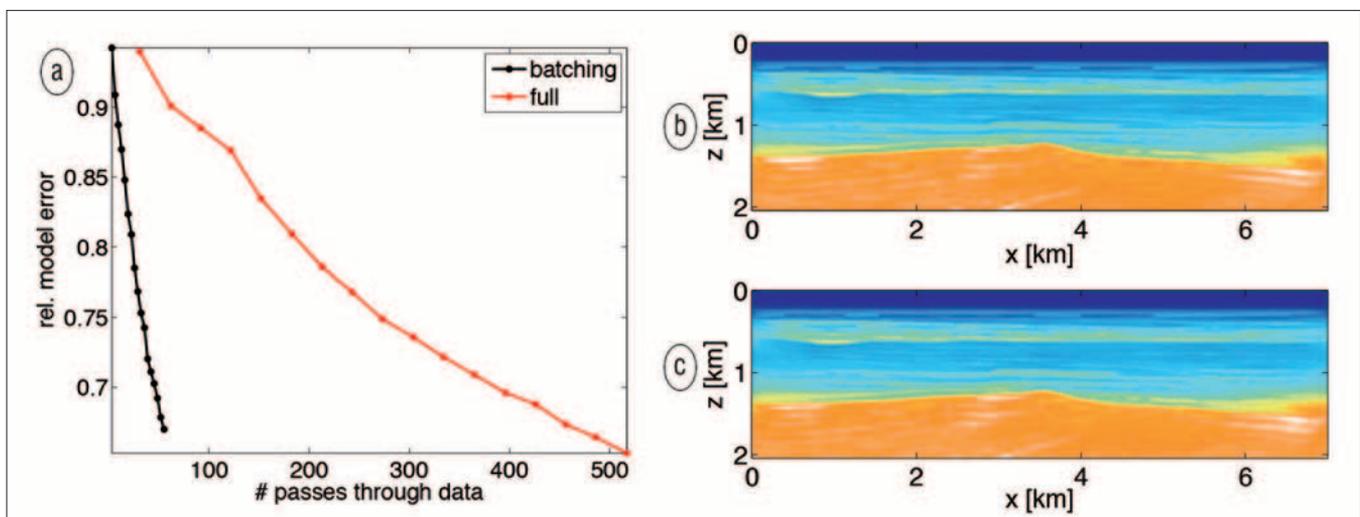
While independent renewals of the source subsets certainly leads to a speedup of the convergence of the algorithm,

choosing these subsets continues to introduce random errors into the optimization, which eventually stalls progress of the algorithm, in particular when there is noise (e.g., Krebs et al., 2009, and van Leeuwen et al., 2011). So the question arises how to choose a dynamic sampling strategy that mitigates the effect of these sampling errors by growing the sample size such that the algorithm continues to converge. The recent theoretical analysis by Aravkin et al. shows that this can be accomplished by increasing the sample size fast enough so that the sampling error decays at least as fast as the inversion makes progress toward the solution. Even though this approach sounds intuitive—why would one need an accurate gradient if one is still far from the solution—it can lead to a somewhat counterintuitive sampling strategy where sample sizes will be chosen to increase slowly only in situations where the optimization converges slowly (e.g., in cases where the problem is ill-posed).

As Figure 2 indicates, significant speedup in the turnaround time of FWI is achievable if we follow the above sampling strategy. However, this ten-fold speedup was obtained in a somewhat ad hoc manner because the convergence rate of FWI is generally not known beforehand and that in turn means that we do not know how to increase the sample size. To address this practical issue, we introduce in the next section a heuristic algorithm that allows us to control errors related to subsampling and forward modeling accuracy dynamically.

**Frugal FWI with adaptive sample-size and simulation accuracy**

Aside from the number of source experiments—whether these are the number of shots or the number of frequencies of a time-harmonic wave-equation simulator—the cost of wave-equation solvers depends on the desired accuracy. By the same token of controlling errors in the sample mean, by increasing the sample size, we can also choose to increase the accuracy of the wave-equation solver as needed by the optimization. For indirect wave-equation solvers, whose accuracy depends on the number of iterations, this leads to



**Figure 2.** (a) Relative error as a function of the number of passes through the data. FWI results on BG Compass (b) passing through all data, and (c) working with subsets reducing the computational cost ten-fold.

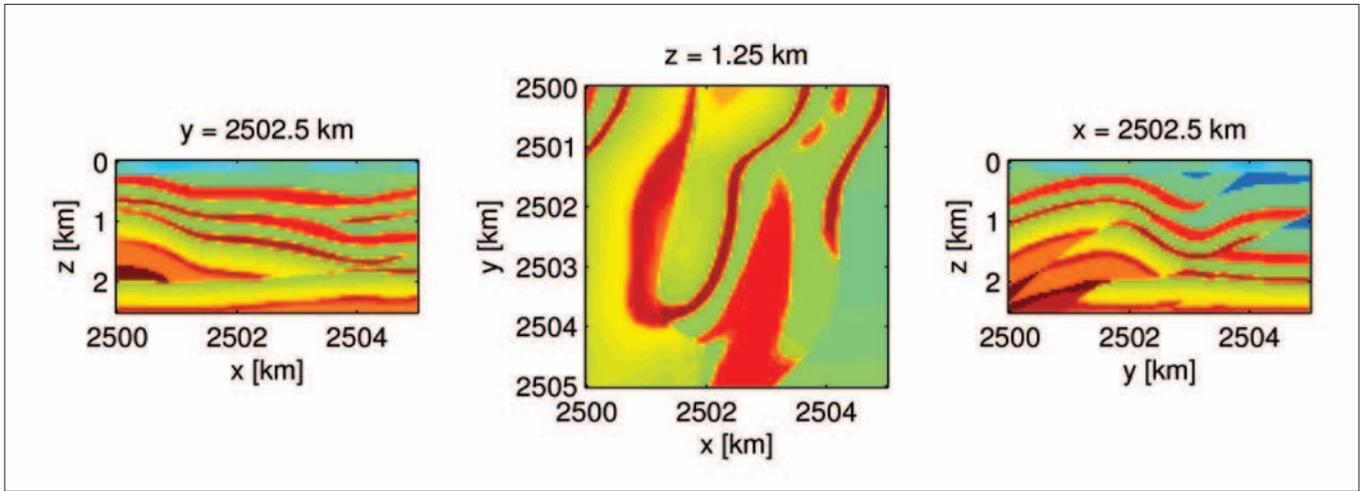


Figure 3. True overthrust velocity model.

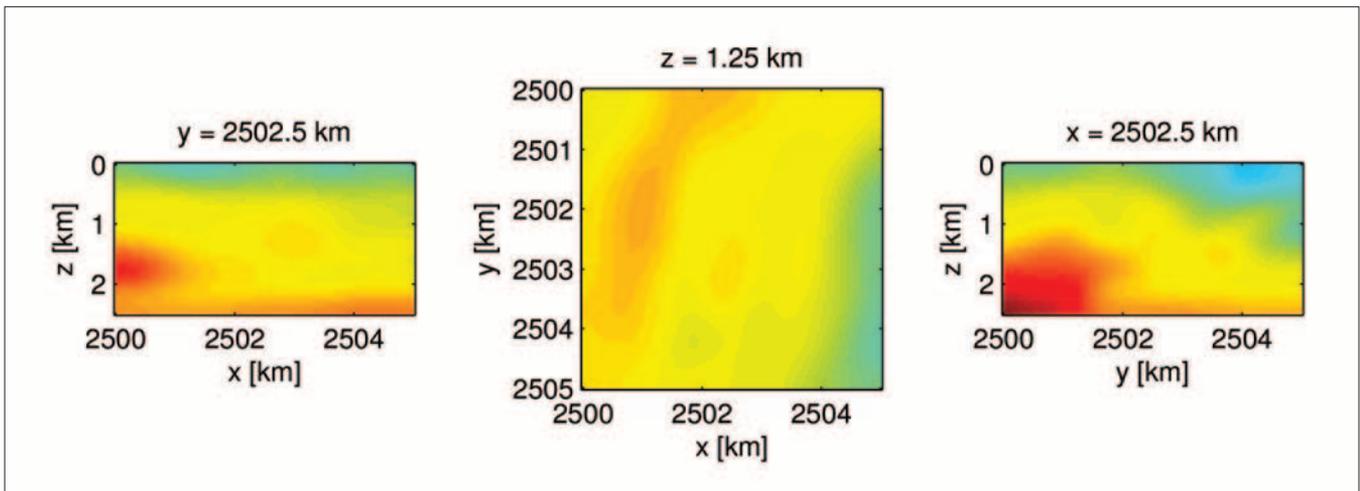


Figure 4. Starting overthrust velocity model obtained by smoothing.

additional savings during the early stages of the inversion where accuracy is not yet needed. However, as before, we will need to control these errors to guarantee convergence. For this purpose, we first show how to calculate the misfit and gradient with selected tolerance.

**Misfit and gradient calculations with selected tolerances**

Key components of our stochastic optimization scheme for FWI are the approximate misfit and gradient calculations with a prescribed but beforehand unknown tolerance. To this end, we introduce a heuristic method that calculates this tolerance because we do not know the true error in the modeling as we do not have access to the true forward modeled result (= exact solution of the wave simulator). Instead, we use an iterative procedure (outlined below or see van Leeuwen and Herrmann, 2013, for further detail) that finds the smallest  $k$  for which

$$|\phi_i(\mathbf{m}, \alpha^k \varepsilon) - \phi_i(\mathbf{m}, \alpha^{k+1} \varepsilon)| \leq \eta \phi_i(\mathbf{m}, \alpha^{k+1} \varepsilon) \quad i \in J,$$

for some  $0 < \alpha < 1$  and for a user defined  $\eta$  and initial tolerance of  $\varepsilon = 10^{-2}$ . During the  $k^{th}$  iteration (with  $k = 1 \dots 10$ )

of this procedure, we solve the wave equation up to  $\varepsilon$  and we compute the  $\ell_2$ -norm of the data residual—i.e., the  $\ell_2$ -norm of the difference between the forward modeled and observed data. If the absolute value of the difference between the norms of the current and previous residual is smaller than an  $\eta$  fraction of the current residual, we are done. Otherwise, we multiply the tolerance  $\varepsilon$  with  $\alpha$  and continue the iterative wave-equation solve with a warm start to reach this new tolerance. We continue to lower the tolerance until the improvement in the residual norm falls below a certain fraction ( $\eta$ ) of the residual norm of the current iteration. Given this empirically found tolerance  $\varepsilon$ , we calculate the adjoint wavefield, the misfit, and the gradient.

The advantage of this approach is that it solves the forward and adjoint wavefields to an accuracy commensurate with the data misfit—i.e., the norm of the residual. This means that, initially, we do not solve the wave equation accurately because the residual is still large at that point because we are still far away from the solution of the FWI problem. As we converge toward the solution, we increase the accuracy, and we do this only when we bring the residue down. In this way, we end up with a highly economic approach to FWI.

**Stochastic FWI with error control**

With the approximate misfit and gradient calculations in place, we are now in a position to put the overall optimization scheme together. For this we need to come up with a scheme that selects the sample size and  $\eta$  parameter, which measures

the fraction to within which we want to reduce the  $\ell_2$ -norm of the residue. We start the procedure by setting the sample size and fraction to some initial value (e.g.,  $b = 1$  and  $\eta = 1$ ), followed by the computation of the misfit function for the starting model  $\mathbf{m}_0$  and for  $b$  randomly selected source(s).

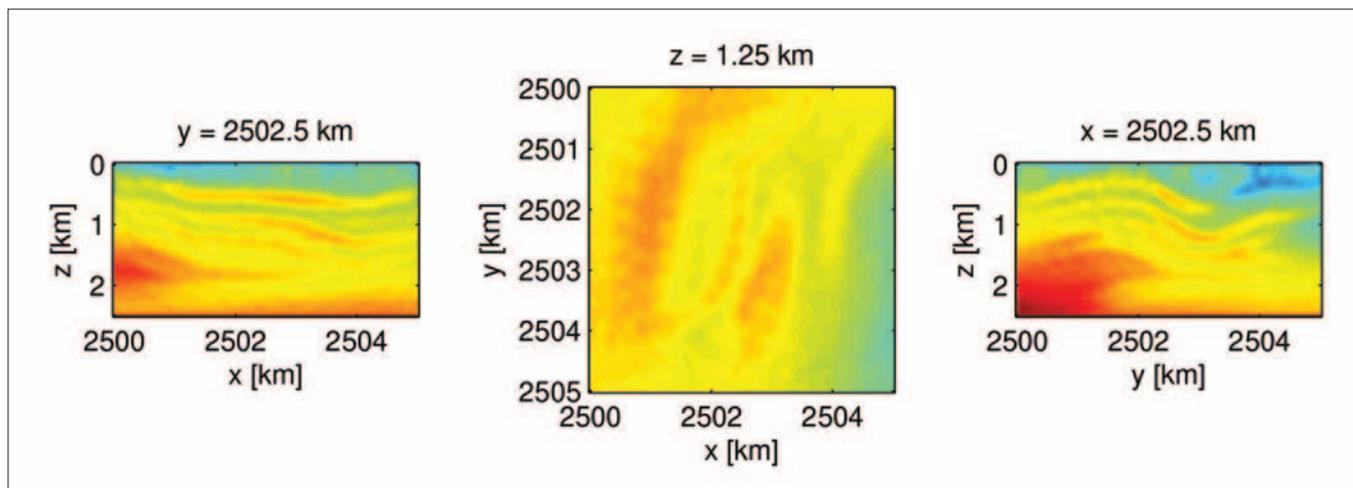


Figure 5. Inversion result by working with all data ( $b = 121$ ).

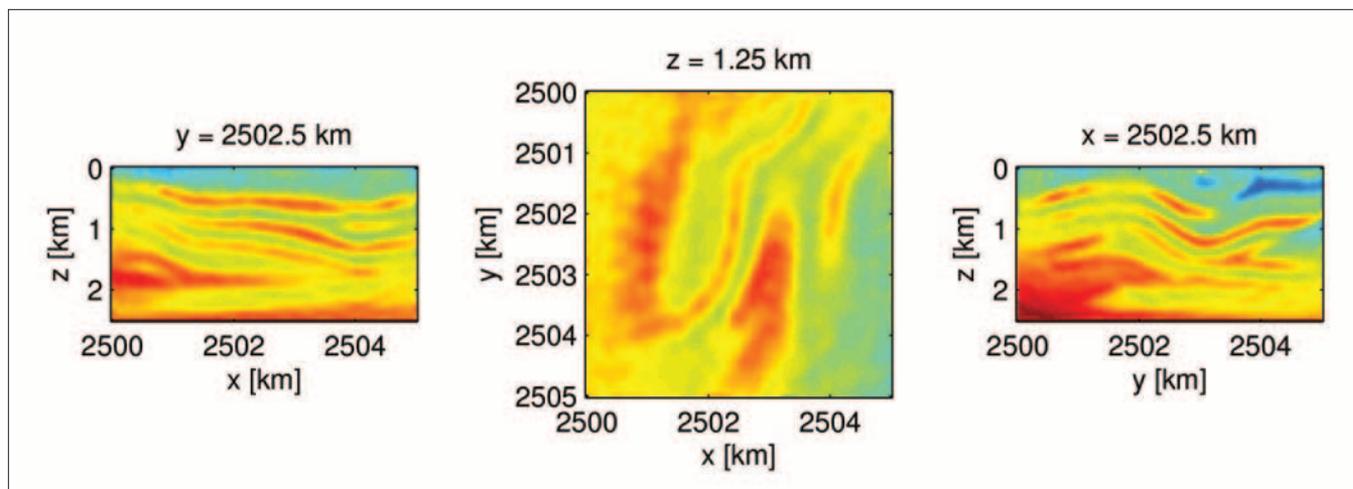


Figure 6. Inversion result by working with one shot ( $b = 1$ ).

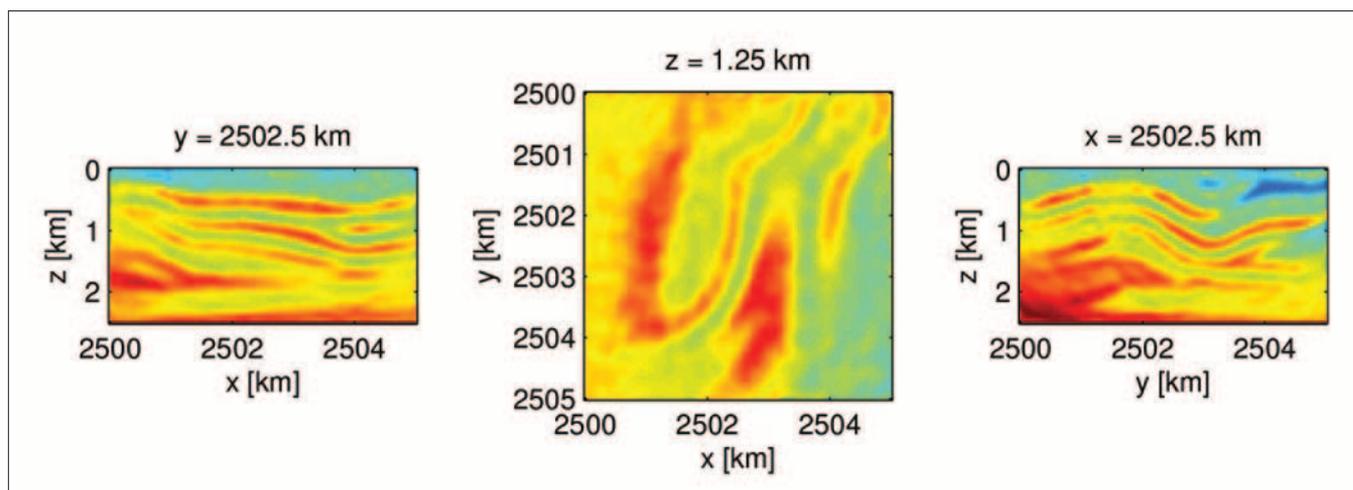


Figure 7. Inversion result for working with increasing sample sizes chosen according to the algorithm described above.

Downloaded 09/27/13 to 134.7.248.132. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

Next, we use the approximated misfit and gradient to compute a linesearch to calculate the correct scaling. If this linesearch is *not* successful, we reduce the fraction  $\eta$  by a factor of two—i.e.,  $\eta \mapsto \eta/2$ . This forces the misfit and gradient to be calculated more accurately. We repeat this procedure until the linesearch becomes successful. At that point, we draw a new subset of random shots and calculate the misfit and gradient for this misfit. Next, we compute the average of the misfits for these two independent subsets and compare this average with the average misfit of the previous iteration. If the average misfit of the current iteration is larger or equal to the average of the previous iteration, we increase the sample size by a preset value (e.g.,  $\beta = 1$  until we reach the total number of source experiments  $M$ ). Otherwise, we compute a linesearch on the last subset and update the model. We repeat this procedure until a maximal number of outer iterations is reached or until the data is fit within some user-defined tolerance.

After each iteration, if the linesearch is successful, we draw a new independent subset of sources and we check for average descent amongst the previous and new source subset selection. If the average misfit does not decrease, the algorithm increases the sample size by  $\beta$ . For a failing linesearch, the algorithm calls for more accuracy for the misfit and gradient calculations by decreasing  $\eta \mapsto \eta/2$ . As a consequence, the algorithm is economic in its use of computational resources because it calls for data and more accuracy only when needed.

This stochastic approach with dynamic sample size and accuracy selection is in sharp contrast with most existing optimization strategies employed by industry that typically insist on working with all data at full accuracy. Among existing economic approaches, there is the fast method with source encoding, advocated by Krebs et al. (2009) and later by van Leeuwen et al. (2011) and Haber et al. (2012), which offer no or practically limited guarantees of convergence. We

overcome this problem by controlling the error by increasing the accuracy and sample size as needed. Before discussing our latest results on a complex 2D synthetic for the Gulf of Mexico using a similar approach, let us first briefly evaluate the performance of our algorithm on a small 3D synthetic.

**Example: Overthrust model**

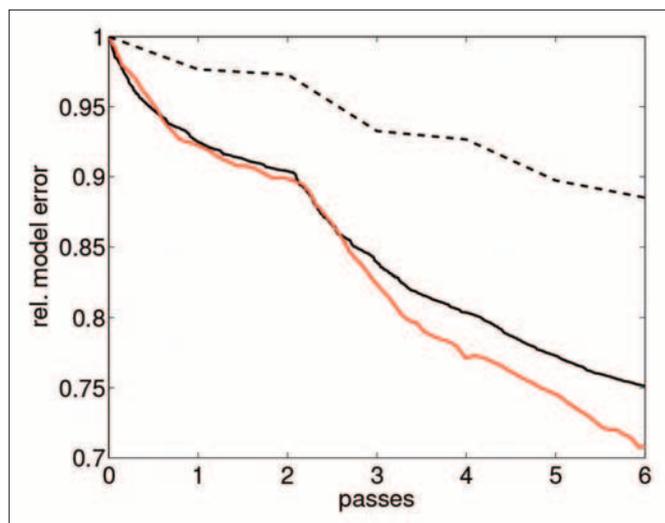
To demonstrate the performance of our proposed frugal approach to FWI, we invert a data set generated with iWave, an open-source time-domain finite-difference code (Terenyev et al.), for a 5-km  $\times$  5-km central part of a well-known benchmark model for seismic inversion (the Overthrust model) with 50-m spacing. The true and initial models for this small example are plotted in Figures 3 and 4, respectively. The data were collected from only 121 sources and 2601 receivers (both regularly spaced).

The inversions are carried out by the above-described algorithms implemented in (parallel) Matlab, with the exception of our row-based preconditioner (van Leeuwen and Herrmann, 2013), which is implemented in C and is called from Matlab using a MEX interface. During the inversion, we use three frequencies  $f = 4, 6,$  and  $8$  Hz and adapt the grid spacing to ensure a minimum of 10 grid points per wavelength. All experiments were conducted on a small Dual-Core SuperMicro system with 2 Intel(R) Xeon(R) CPU E5-2670 0 @ 2.60 GHz and with 128 GB RAM.

To mimic actual real-life situations, where computational resources and turn-around times are constraints, we limit the total passes through the data per frequency to only two and compare the results to the following three scenarios: work with one shot at the time ( $b = 1$ ), work with all shots ( $b = 121$ ), or choose the number of shots adaptively. In all cases, we select the accuracy of the simulations as needed. While the amount of work in these three scenarios is the same, we would expect the results for the first and third method to be superior compared to the scenario where we insist to work with all data for each misfit and gradient calculation. Indeed, Figures 5 and 7 confirm this prediction and shows significant improvements in the inversion results when we work with subsets of the data only. This improvement is also reflected in the plots for the relative model error, which we included in Figure 8. These plots show that working with subsets pays off when one has access only to a limited computational budget. While the improvement of our dynamic method may not be drastic, compared to the method working on single randomly selected shot for each iteration (compare Figures 6 and 7), we have guarantees that our dynamic method will likely converge.

**Case study: Chevron Gulf of Mexico data set**

Aside from reducing the demand on computational resources, the presented randomized source subsampling technique also creates the possibility to use sophisticated regularization techniques to stabilize FWI for complex geological areas and for situations where the simulations do not totally capture the physics of the problem, e.g., by inverting elastic data using a scalar velocity-only acoustic code. The Chevron Gulf of Mexico data set is an example of such a situation where



**Figure 8.** Relative model errors as a function of the number of passes through the data (two per frequency  $\times$  three frequencies). The black dashed line denotes the model error when we work with all data while the solid black and red lines denote the errors when working with a single shot only or dynamically.

Downloaded 09/27/13 to 134.7.248.132. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

the presence of salt and elastic phases, “lack of long-enough offsets,” and poor signal to noise at low frequencies make for a challenging data set for acoustic-only FWI. To overcome these realistic but highly disadvantageous circumstances where turning waves do not make it below the salt, we developed the following workflow:

- *Starting model building.* To get a reasonable starting model, we conducted first-break traveltimes inversion on hand-picked traveltimes from approximately 600,000 traces. We used the relatively noise-free high-frequency arrivals for this purpose. With these picks, we carried out a traveltimes inversion, yielding an rms traveltimes misfit of only 11 ms. The result of this exercise, including the raypaths, is plotted in Figure 9. The lack of deep penetration of the rays clearly reveals the challenges we face for the recovery of the salt bodies. The hope of FWI is to improve the recovery of top and bottom salt and hence improve the quality of the subsalt imaging. We want to accomplish this with minimal user intervention.
- *Denoising.* To overcome the extremely poor signal-to-noise at low frequencies, we carried out curvelet-domain denoising on selected frequency slices (2–5 Hz) in the source-receiver domain. The denoising itself, consisted of solving a sparsity-promoting program to find the denoised curvelet-coefficients, followed by a debiasing step and the inverse curvelet transform, to recover the denoised frequency slices. An example of this sophisticated denoising procedure is included in Figure 10. This result clearly shows that we are able to remove a significant portion of the incoherent noise while preserving the amplitudes (there is little coherent energy in the difference plot). Note that the 2D curvelet transform decomposes a 2D function into localized plane waves that live in different frequency bands. Curvelets are sparse (i.e., there are only a few large coefficients) because they give rise to significant coefficients only if the orientation and dominant bandwidth of the curvelet coincide with a localized event in the data. This makes this transform suitable for denoising.
- *FWI with nonlinear smoothing of the model updates.* Given the denoised frequency slices, we ran our inversion in seven overlapping low-to-high frequency bands with four randomly selected frequencies per band. To help stabilize this multiscale inversion, we apply a nonlinear anisotropic smoothing by imposing a sparsity constraint on the curvelet-transform coefficients of the model updates, as described by Li et al. (2012). Because imposing this constraint calls for several iterations, we follow the frugal strategy of this paper by working with only 600 randomly selected shots out of a total of 3200 shots. The results of this exercise for six Gauss-Newton subproblems is included in Figure 11.

The output of this workflow is encouraging because we are evidently able to recover top salt with reasonable detail without the need of (1) reflection traveltimes picking and salt flooding that are part of the conventional velocity model building and migration workflows and which typically require large amounts of human intervention, and (2) data (pre)processing

prior or during the inversion to remove the imprint of elastic phases in the data.

The latter feature is remarkable because we are able to get a reasonable inversion result for shallow depths without manipulating the data (except for the denoising of course). Because we are carrying out the inversion of the elastic data with an acoustic code, we had to limit the number of subproblems to prevent the algorithm from fitting the elastic phases. This also explains that we were able to fit only 21% of the data's energy and that our result lacks details of the top salt

While these results may be encouraging, except of course for the fact that we had to pick first-break traveltimes from some 600,000 traces, there is still room for improvement, including:

- Obtaining better starting models or less sensitivity to the quality of the initial model. While first-break traveltimes inversion gave us a reasonable starting model for shallow depths, it failed to provide sufficient accurate velocity information to recover the long-wavelength components of the salt. Because the offsets were “limited” to 20 km and there is no reasonable data quality below 2 Hz, FWI is not able to recover the bottom of the salt. Turning waves simply do not make it under the salt for these offset and frequency ranges.
- Incorporation of more wave physics to explain the elastic phases in the data.

## Discussion

We presented a formulation of FWI that is cognizant of its often excessive and sometimes even unfeasible demand for computational resources by calling for more data and accuracy of wave simulations only as needed to guarantee convergence of the algorithm. This approach allowed us to conduct 3D FWI on a small machine. The same data reduction techniques enabled us to regularize FWI via nonlinear anisotropic curvelet-domain smoothing of the model updates. Application of this approach to the Chevron GOM data set yielded reasonable results without extensive user intervention (except, of course, for the first-break picking to estimate the starting model). However, at depths exceeding 4 km, the starting model lacked information on the long-wavelength components of the salt. Because of the relatively short offset and missing low frequencies, FWI was unfortunately not able to recover these missing long-wavelength components of the salt.

Aside from the somewhat obvious calls for improvements resulting from our work on the GOM data set, FWI more generally suffers fundamentally from lack of robustness with respect to outliers (e.g., unmodeled elastic phases in the data) and worse from local minima stemming from the highly nonlinear relationship between the medium properties and the data. Because “getting the physics right” really is an oxymoron, FWI requires a formulation that is less sensitive to noise, outliers in the data, and scaling by the source function. Early attempts toward such a formulation have appeared in the literature in recent work by Aravkin et al. on FWI with students'  $\mathbf{t}$  penalties and by the same authors on the estimation of nuisance parameter with variable projection, which includes source estimation. This work also

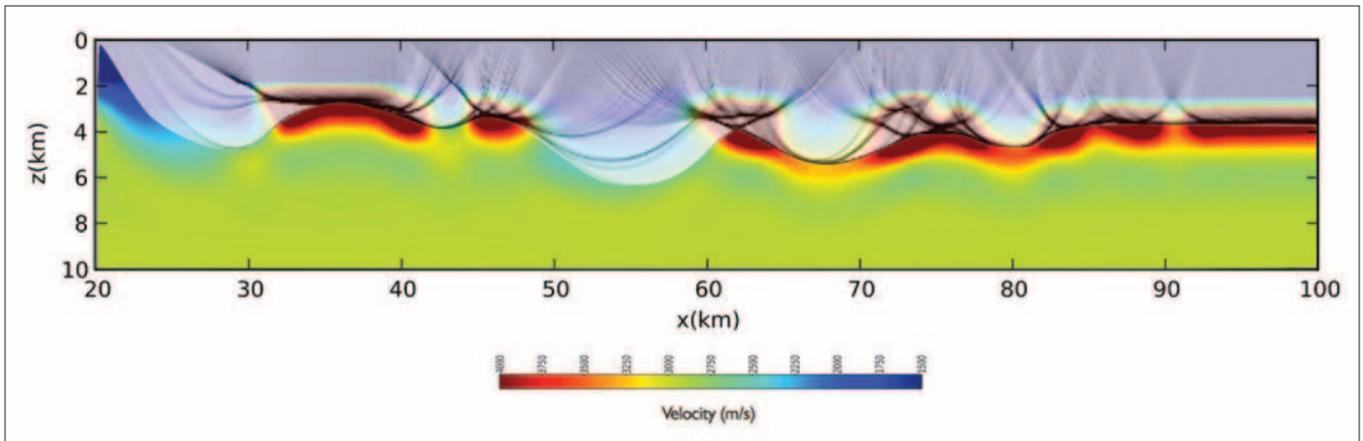


Figure 9. Initial model including rays obtained by first-break traveltime picking.

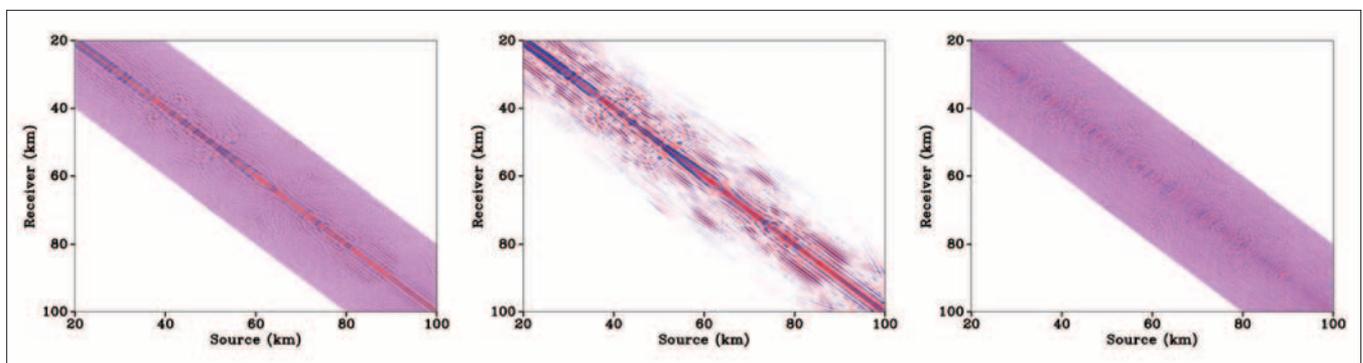


Figure 10. Curvelet denoising of the real part frequency slice at 2 Hz; (Left) noisy input data, (middle) denoised data, and (right) difference plot.

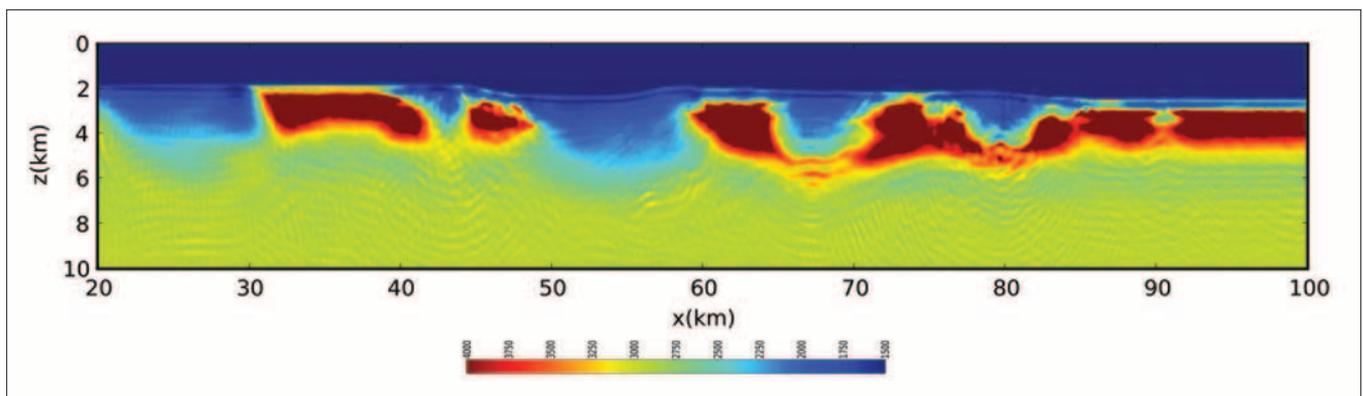


Figure 11. Final result of the curvelet sparsity constrained FWI on the GOM data set.

showed that penalty functions derived from the students'  $t$  distribution can, contrary to  $\ell_2$ - and  $\ell_1$ -norm misfits, handle outliers such as unmodeled bad traces and elastic phases. As for the fundamental problem of mitigating the effects of local minima, several approaches have been proposed recently, including the following:

- a rigorous mathematical framework with frequency selection that guarantees convergence of steepest descent techniques by a multilevel method that relaxes the parsimoniousness of the model as the frequency increases. This theoretical work by de Hoop and coworkers has connections to the curvelet regularization we used to invert

the GOM data. Unfortunately, this work merely guarantees convergence for certain starting models and does not remove the condition that the starting model needs to be accurate enough to fall in the region of convergence;

- extension of FWI to include focusing. This approach, advocated by Symes and Biondi, seeks to minimize the difference between the observed and simulated data while also minimizing a MVA-like objective. Aside from including the principle of focusing, this approach aims to remove local minima by increasing the degrees of freedom. While this leads to a computationally challenging formulation (see recent work by the authors for a possible way around this computational problem), there are indications that

the larger degree of freedom partly eliminates the need of a highly accurate starting model, and

- removal of the nonlinearity of FWI by working with an “all-at-once” approach where the optimization is carried out over the wavefields as well as over the model parameters. Because of the increase of the degrees of freedom, we can also expect this approach to be less sensitive to local minima as was confirmed in recent work by the authors. This work proposes a new penalty method, which overcomes the infeasible storage demands of the all-at-once method. **TLE**

## References

- Aravkin, A., M. Friedlander, F. Herrmann, and T. van Leeuwen, 2012, Robust inversion, dimensionality reduction, and randomized sampling: *Mathematical Programming*, **134**, no. 1, 101–125, <http://dx.doi.org/10.1007/s10107-012-0571-6>.
- Haber, E., M. Chung, and F. Herrmann, 2012, An effective method for parameter estimation with PDE constraints with multiple right-hand sides: *SIAM Journal on Optimization*, **22**, no. 3, 739–757, <http://dx.doi.org/10.1137/11081126X>.
- Krebs, J., J. Anderson, D. Hinkley, A. Baumstein, S. Lee, R. Neelamani, and M.-D. Lacasse, 2009, Fast full wave seismic inversion using source encoding: 79th Annual International Meeting, SEG, Expanded Abstracts, 2273–2277, <http://dx.doi.org/10.1190/1.3255314>.
- van Leeuwen, T., A. Aravkin, and F. Herrmann, 2011, Seismic waveform inversion by stochastic optimization: *International Journal of Geophysics*, <http://dx.doi.org/1155/2011/889041>.
- van Leeuwen, T. and F. Herrmann, 2012, Fast waveform inversion without source-encoding: *Geophysical Prospecting*, <http://dx.doi.org/10.1111/j.1365-2478.2012.01096.x>.
- van Leeuwen, T. and F. Herrman, 2013, 3D frequency-domain seismic inversion with controlled sloppiness, UBC, <https://www.slim.eos.ubc.ca/Publications/Private/Tech>.
- van Leeuwen, T. and F. Herrman, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: *Geophysical Journal International*, first published online July 30, 2013, <http://dx.doi.org/10.1093/gji/ggt258>.
- Li, X., A. Aravkin, T. van Leeuwen, and F. Herrmann, 2012, Fast randomized full-waveform inversion with compressive sensing: *Geophysics*, **77**, no. 3, A13–A17, <http://dx.doi.org/10.1190/geo2011-0410.1>.
- Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: *Geophysical Journal International*, **167**, no. 2, 495–503, <http://dx.doi.org/10.1111/j.1365-246X.2006.02978.x>.
- Tarantola, A. and A. Valette, 1982, Generalized nonlinear inverse problems solved using the least squares criterion: *Reviews of Geophysics and Space Physics*, **20**, no. 2, 219–232, <http://dx.doi.org/10.1029/RG020i002p00219>.

*Acknowledgments: This work was financially supported by the Natural Sciences and Engineering Research Council of Canada Collaborative Research and Development Grant DNOISE II (375142-08) and was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, BP Global, CGG, Chevron Corporation, ConocoPhillips, ION GXT, Petrobras, PGS, Total SA, WesternGeco, and Woodside Energy Ltd.*

*Corresponding author: fherrmann@eos.ubc.ca*