On the choice of material parameters for elastic waveform inversion

Ryan Modrak, Yanhua Yuan and Jeroen Tromp, Princeton University*

SUMMARY

The material parameterization used to compute derivatives, search directions, and model updates in elastic waveform inversion can have a significant effect on the robustness and efficiency of the overall nonlinear optimization procedure. For isotropic media, conventional wisdom holds that it is better to work with compressional- and shear-wave speeds α and β than Lamé parameters λ and μ. Conventional wisdom further holds that, given their improved scaling and reduced covariance, bulk and shear moduli κ and μ, or "wave speed-like" parameters $\sqrt{\frac{\kappa}{\rho}}$ and $\sqrt{\frac{\mu}{\rho}}$, are an even better choice. Numerical tests involving hundreds of inversions, six subsurfaces models, and three full-waveform misfit functions, however, reveal a more complicated picture.

Working in the time domain and using numerically safeguarded quasi-Newton model updates, we find that the relative performance of material parameterizations is strongly problem dependent. Despite their physical relevance, α and β rank last among all parameterizations in both efficiency and robustness. In a large number of cases, λ and μ outperform all competitors. These results make sense in light of some additional observations, namely, that (1) relative performance of material parameterizations correlates with nonlinearity, which can be affordably quantified in terms of deviation of the misfit function from a quadratic form, and (2) wave-speed like parameters perform well for phase-based inversions, but not for waveform- or envelope-based inversions. Since a direct relation between wave-speed perturbations and data residuals exists for phase misfit functions, but not for envelope- or waveform-difference misfit, good performance of Lamé parameters relative to wave speed-like parameters in the latter cases is perhaps not surprising.

INTRODUCTION

The choice of one set of material parameters or another can be regarded as a numerical decision, since it affects the conditioning of an inversion without any change to the underlying misfit function. Because poor numerical conditioning can lead to global convergence difficulties, the material parameterization has a bearing not only on computational efficiency, but also on the ultimate success or failure of an inversion. Such facts were recognized early on in geophysics and other computational fields (e.g., Tarantola 1986, Cai and Keyes 2002).

In one of the earliest and most influential elastic waveform inversion studies, Tarantola (1986) considered the choice of material parameters and recommended compressional- and shear-wave speed over Lamé parameters. This recommendation, it appears, was based mostly on analysis of diffraction pattern responses to different material perturbations; other types of experiments are mentioned, but only in passing. In later work, Tarantola advocated bulk and shear moduli, which display an improved scaling and reduced covariance relative to α and β.

While the diffraction pattern analysis Tarantola developed has proven useful, it seems unlikely he would have relied so heavily on the device today, given the vastly greater computational resources available. "Brute force" experiments involving hundreds of inversions now provide a feasible, and we argue, much more useful method for comparing material parameterizations.

THEORY

The gradient of a full-waveform misfit function in an isotropic elastic inversion can be written as

$$\delta\chi = \int K_\rho \delta \ln\rho + K_\kappa \delta \ln\kappa + K_\mu \delta \ln\mu \; \mathrm{d}x\,, \tag{1}$$

where ρ, κ, μ are density, bulk modulus and shear modulus, respectively, and K_ρ, K_κ, K_μ are kernels relating variations in the above material parameters to data misfit. Formulas for the kernels

$$K_\rho(x) = \int \rho(x)\, \partial_t s(x,t) \cdot \partial_t s^\dagger(x,t)\, \mathrm{d}t \tag{2}$$

$$K_\kappa(x) = \int \kappa(x)\, \nabla\cdot s(x,t)\, \nabla\cdot s^\dagger(x,t)\, \mathrm{d}t \tag{3}$$

$$K_\mu(x) = \int 2\mu(x)\, D(x,t) : D^\dagger(x,t)\, \mathrm{d}t \tag{4}$$

resemble classical imaging conditions, involving interaction between forward and adjoint wavefields s and $s^\dagger$ or associated deviatoric strain tensors D and $D^\dagger$ (Tromp et al., 2005).

Of course, $\delta\chi$ can be expressed in other ways as well. For example, kernels for Lamé parameters corresponding to $\delta\lambda$ and $\delta\mu$ with density held fixed are given by

$$K_\lambda = \big(1 - \tfrac{2\mu}{3\kappa}\big) K_\kappa \tag{5}$$

$$K'_\mu = \tfrac{2\mu}{3\kappa} K_\kappa + K_\mu\,. \tag{6}$$

Similarly, kernels for compressional- and shear-wave speed with fixed density are

$$K_\alpha = 2\big(1 + \tfrac{4\mu}{3\kappa}\big) K_\kappa \tag{7}$$

$$K_\beta = 2\big(K_\mu - \tfrac{4\mu}{3\kappa} K_\kappa\big)\,, \tag{8}$$

and kernels for bulk-sound speed (i.e., $\phi = \sqrt{\frac{\kappa}{\rho}}$) and shear-wave speed with fixed density are

$$K_\phi = 2K_\kappa \tag{9}$$

$$K'_\beta = 2K_\mu\,. \tag{10}$$

More details concerning eqs. 1–8 can be found in Tromp et al. (2005), Zhu et al. (2009) and Luo et al. (2013).

TEST PROBLEMS

To compare material parameterizations, we ran inversions with the following target models: (a) *Marmousi onshore*, (b) *Marmousi offshore*, (c) *overthrust onshore*, (d) *overthrust offshore*, (e) *BP 2007 diapir*, (f) *BP 2007 anticline*. For (a) and (c) we used the standard IFP Marmousi and SEG/EAGE overthrust models, and for (b) and (d) we added a 500 m water layer. For (e) and (f) we cropped the full BP 2007 model between $x = 20$–45 km and $x = 37.5$–62.5 km, respectively, and down-scaled in both x and z by a factor of two. Shear-wave speeds were obtained from compressional-wave speeds by $\beta = \frac{1}{\sqrt{3}}\alpha$, and densities were obtained using Gardner's law $\rho = 0.31\alpha^{0.25}$. Starting models were derived from target models by convolving with a Gaussian of standard deviation 600 m.

We carried out inversions with waveform-, envelope- and phase-difference misfit functions given by

$$\chi_1(m) = \sum \int |s(m,t) - d(t)|^2 \, \mathrm{d}t \tag{11}$$

$$\chi_2(m) = \sum \int |E_s(m,t) - E_d(t)|^2 \, \mathrm{d}t \tag{12}$$

$$\chi_3(m) = \sum \int |\phi_s(m,t) - \phi_d(t)|^2 \, \mathrm{d}t \,, \tag{13}$$

where s are synthetics, d are data, E is envelope, ϕ is phase, and the sum is taken over all sources, receivers, and components. Rather than instantaneous phase (e.g., Bozdag et al. 2011), we defined ϕ as the analytic signal divided by the envelope. This new phased-based misfit function is effective, we find, in reducing cycle-skipping artifacts and probably deserves a more detailed examination like that given to χ_2 by Yuan et al. (2015).

Borrowing a device from the numerical optimization literature, we measured the nonlinearity of each model/misfit combination in terms of the size of the third derivatives of the misfit function (Nash and Nocedal, 1991). Because the third derivatives vanish if the misfit function is quadratic, the expression

$$\mathrm{DQ} = \frac{\|g|_{m_\mathrm{init}} - g|_{m_\mathrm{true}} - H|_{m_\mathrm{true}} p\|_\infty}{\|p\|_\infty^2} \,, \tag{14}$$

where g is the gradient, H is the Hessian and $p = m_\mathrm{init} - m_\mathrm{true}$, provides a measure of deviation from a quadratic form. We find that the resulting ranking, shown in Table 1, agrees with our subjective sense of how easy or hard each test problem is (with the somewhat odd exception of the BP 2007 anticline test case, which has an unexpectedly high DQ).

TESTING PROCEDURES

For each model/misfit/material combination, we carried out a suite of inversions, each with the same starting model but with different frequency bands. While varying the frequency band between inversion, we kept the dominant frequency fixed within each inversion, forgoing any type of multiscale procedure in the manner of Bunks et al. (1995).

In these experiments, forward and adjoint simulations were carried out using `SPECFEM2D`, with full elastic/acoustic coupling for offshore problems (Komatitsch and Vilotte, 1998; Luo et al., 2013). Nonlinear optimization, data preprocessing, gradient postprocessing, and workflow integration tasks were performed in the `SeisFlows` framework (http://github.com/PrincetonUniversity/seisflows). In generating data and synthetics, an absorbing boundary condition was used to exclude multiple reflections.

To avoid problematic tradeoffs between parameters, we treated density as a dependent variable through Gardner's relation. While such relations can be used to modify the gradient directly, i.e., by substituting $\delta\rho = \frac{\partial\rho}{\partial\kappa}\delta\kappa + \frac{\partial\rho}{\partial\mu}\delta\mu$ in eq. 1, a more effective approach, we find, is to simply drop the $K_\rho\delta\rho$ term, that is, to update all parameters aside from density by their respective kernels and then, using these new values, update density by its empirical relation.

To avoid complicated choices about penalty function weights, we employed a simple "regularization by convolution" approach in which kernels were convolved with a Gaussian with a fixed standard deviation of 5 numerical grid points, or roughly 100 m (e.g., Tape et al. 2007).

To ensure meaningful comparisons between inversion results, close attention was paid to the nonlinear optimization procedure. Inversions were run using L-BFGS with a memory of five previous gradient evaluations and with M3 scaling (Liu and Nocedal, 1989), stopping after convergence to a minimum of the misfit function or 100 model updates, whichever occurred first. For comparison, a subset of inversions was rerun using truncated Newton model updates with Eisenstat-Walker stopping condition and L-BFGS preconditioner, stopping after convergence or 50 model updates. A backtracking line search with bound constraints was used with both types of search directions (Dennis and Schnabel, 1996).

Numerical safeguards are important in L-BFGS inversions because the accuracy of the quasi-Newton approximation to the inverse Hessian is known to break down on occasion. Besides line search bounds, restarting the nonlinear optimization algorithm in the manner of Powell (1977) is essential. By restarting when the angle between the steepest descent direction and search direction exceeds a certain threshold, say 85 degrees, we adapt Powell's restart condition to L-BFGS. Restarting is usually effective, we find, in restoring fast convergence if an inversion becomes stalled. Details about the number of restarts in the waveform-difference (χ_1) inversions are given in Table 2. Compared with acoustic inversion, the need for restarting in elastic inversion is much greater.

To rate the performance of an inversion, we used the following measure of error reduction

$$\Delta E = \sum_{i=1}^{N} w_i \left(\|m_{i-1} - m_\mathrm{true}\| - \|m_i - m_\mathrm{true}\| \right), \tag{15}$$

where N is the total number of model updates. Unlike the choice of parameterization to compute model updates with, the choice of one parameterization or another in eq. 15 is not especially significant. To permit comparisons, we converted other

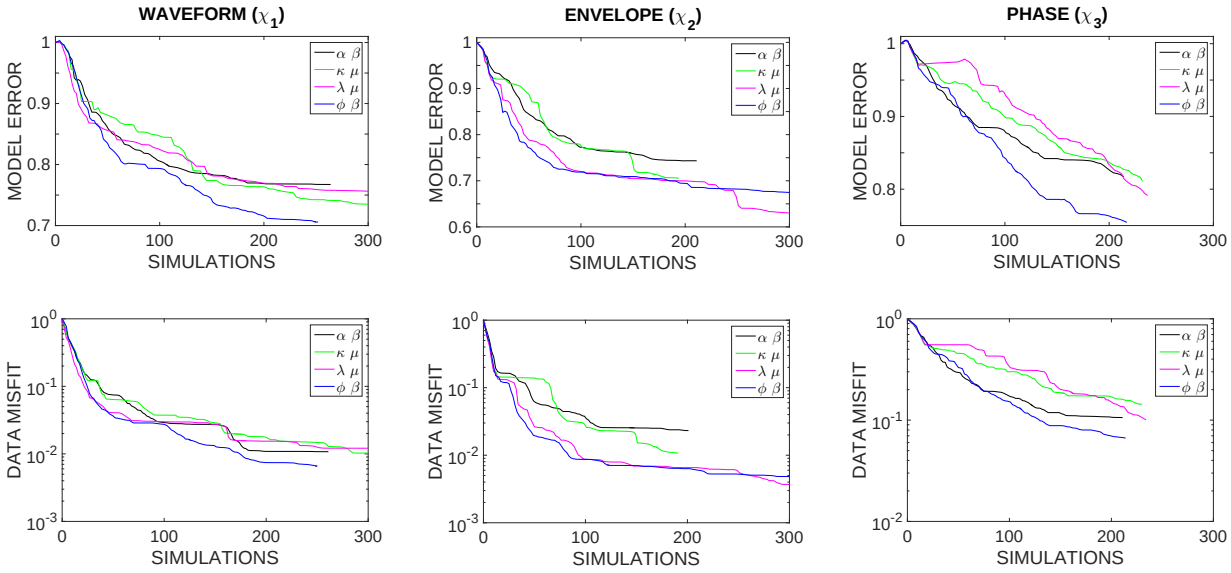

Figure 1: Model and data space convergence plots for the *Marmousi onshore 3 Hz* test case quasi-Newton inversions. In as far as phase misfit tends to perform well with ϕ, β and envelope misfit tends to perform well with λ, μ, these results are representative of the entire suite of inversions. By "simulations" we mean the cumulative number of forward and adjoint simulations normalized by the number of sources, or equivalently, the cumulative number of function and gradient evaluations. While in this example model error closely tracks data misfit, behavior in the two spaces is not always so strongly correlated.

parameter sets to α and β and used $\|m\| = \|m_\alpha\|_2 + \|m_\beta\|_2$ in calculating ΔE, where $\|\cdot\|_2$ is the L_2 norm. Letting F_i be the cumulative number of function and gradient evaluations through the i-th model update, our choice of weights

$$w_i = \left(F_i - F_{i-1}\right)^{-1} 2^{-0.01F_i} \tag{16}$$

rewards progress early in an inversion over progress later on. Such weights are important because without them eq. 15 would reduce to $\|m_0 - m_{\text{true}}\| - \|m_N - m_{\text{true}}\|$. It follows that some type of weighting is required to distinguish between two inversions that converge to the same result at different computational expense.

RESULTS

Tables 3 and 4 list which parameter sets, of the four considered, performed best and worst in the quasi-Newton inversions in terms of ΔE. Asterisks denote near ties—in which the winner came within 2.5 percent of the next-closest competitor. For illustration, convergence plots for the Marmousi onshore 3 Hz quasi-Newton inversions are shown in Figure 1. Plots for all other test cases are available online (http://github.com/rmodrak/SEG_2016_abstract).

In the results, strong problem dependence is evident with major differences between onshore and offshore test cases related to underlying differences in nonlinearity. Strikingly, λ and μ ranked first in more than 30 percent and α and β ranked last in more than 65 percent of test cases—despite the latter's physical relevance and the former's lack thereof. Besides poor scaling and large covariance between α and β, much of this behavior, we believe, can be explained in terms of relationships between material parameters and data misfit—in particular, the fact that wave speed perturbations are more directly related to phase differences and than waveform or envelope differences.

To check whether our rankings depended strongly on any aspect of our L-BFGS implementation, we compared quasi-Newton with truncated Newton inversion results. Truncated Newton was less computationally efficient than L-BFGS, which is not inconsistent with Métivier et al. (2014) and not unexpected given the lack of sophisticated preconditioning (Akçelik et al., 2002). Although inversions based on α and β did tend to fail less often with truncated Newton model updates, rankings were not significantly affected. Leaving aside computational cost differences, the similarity between L-BFGS and truncated Newton results, in our view, suggests that robustness is more a matter of numerical safeguards than a question of one nonlinear optimization algorithm over another.

CONCLUSIONS

Results of "brute force" experiments carried out with attention to numerical details demonstrate that the choice of material parameters in elastic waveform inversion is, at the very least, more complicated than the current literature reflects. Additional tests are needed to explore why, despite theoretical considerations that seem to suggest otherwise, Lamé parameters are an effective choice, and to consider issues such as noisy data, frequency- versus time-domain implementation, and implications for anisotropic problems. Whatever the ultimate findings, it seems clear that, in addition to physical or seismological insights, numerical insights are needed for progress in elastic waveform inversion.

Problem	DQ
Marmousi onshore 3 Hz	10^5
Marmousi onshore 4 Hz	10^6
Marmousi onshore 5 Hz	10^6
Marmousi onshore 6 Hz	10^6
Marmousi onshore 7 Hz	10^7
overthrust onshore 3 Hz	10^6
overthrust onshore 4 Hz	10^6
overthrust onshore 5 Hz	10^7
overthrust onshore 6 Hz	10^7
overthrust onshore 7 Hz	10^8
overthrust offshore 3 Hz	10^{11}
overthrust offshore 4 Hz	10^{12}
overthrust offshore 5 Hz	10^{12}
overthrust offshore 6 Hz	10^{12}
overthrust offshore 7 Hz	10^{12}
BP 2007 diapir 3 Hz	10^{12}
BP 2007 diapir 4 Hz	10^{12}
BP 2007 anticline 3 Hz	10^{12}
BP 2007 anticline 4 Hz	10^{13}
BP 2007 anticline 5 Hz	10^{13}
Marmousi offshore 3 Hz	10^{12}
Marmousi offshore 4 Hz	10^{13}

Table 1: Deviation from quadratic. The above ranking, based on waveform-difference misfit (χ_1) agrees mostly with our subjective sense of difficulty.

Problem	α,β	κ,μ	λ,μ	ϕ,β
Marmousi onshore 3 Hz	4	8	10	2
Marmousi onshore 4 Hz	3	11	12	0
Marmousi onshore 5 Hz	1	7	7	0
Marmousi onshore 6 Hz	1	8	6	0
Marmousi onshore 7 Hz	0	4	5	2
overthrust onshore 3 Hz	2	6	4	1
overthrust onshore 4 Hz	3	7	7	2
overthrust onshore 5 Hz	2	6	6	2
overthrust onshore 6 Hz	1	5	7	2
overthrust onshore 7 Hz	3	6	6	4
overthrust offshore 3 Hz	2	2	4	2
overthrust offshore 4 Hz	1	3	5	2
overthrust offshore 5 Hz	1	2	2	0
overthrust offshore 6 Hz	2	3	1	2
overthrust offshore 7 Hz	0	3	4	3
BP 2007 diapir 3 Hz	0	33	7	1
BP 2007 diapir 4 Hz	2	7	8	5
BP 2007 anticline 3 Hz	0	1	1	1
BP 2007 anticline 4 Hz	1	1	1	1
BP 2007 anticline 5 Hz	1	0	0	0
Marmousi offshore 3 Hz	0	4	3	3
Marmousi offshore 4 Hz	1	5	4	0

Table 2: Number of restarts in the waveform-based (χ_1) inversions. A high number of restarts, we find, does not necessarily indicate a given parameterization is performing badly.

Problem	Envelope	Waveform	Phase
Marmousi onshore 3 Hz	λ,μ	ϕ,β	ϕ,β
Marmousi onshore 4 Hz	ϕ,β	ϕ,β	ϕ,β
Marmousi onshore 5 Hz	ϕ,β	ϕ,β	—
Marmousi onshore 6 Hz	ϕ,β	ϕ,β	—
Marmousi onshore 7 Hz	ϕ,β	ϕ,β	—
overthrust onshore 3 Hz	λ,μ	$\lambda,\mu*$	$\lambda,\mu*$
overthrust onshore 4 Hz	λ,μ	ϕ,β	ϕ,β
overthrust onshore 5 Hz	λ,μ	ϕ,β	$\phi,\beta*$
overthrust onshore 6 Hz	λ,μ	$\phi,\beta*$	ϕ,β
overthrust onshore 7 Hz	$\kappa,\mu*$	$\phi,\beta*$	$\phi,\beta*$
overthrust offshore 3 Hz	ϕ,β	ϕ,β	ϕ,β
overthrust offshore 4 Hz	—	λ,μ	ϕ,β
overthrust offshore 5 Hz	—	λ,μ	—
overthrust offshore 6 Hz	—	λ,μ	—
overthrust offshore 7 Hz	—	λ,μ	—
BP 2007 diapir 3 Hz	—	α,β	ϕ,β
BP 2007 diapir 4 Hz	—	κ,μ	—
BP 2007 anticline 3 Hz	—	λ,μ	κ,μ
BP 2007 anticline 4 Hz	—	λ,μ	κ,μ
BP 2007 anticline 5 Hz	—	λ,μ	—
Marmousi offshore 3 Hz	—	α,β	κ,μ
Marmousi offshore 4 Hz	—	ϕ,β	—

Table 3: Best-performing material parameterization in terms of model error reduction (ΔE). Asterisk indicates a near tie. Dash denotes all parameterizations failed.

Problem	Envelope	Waveform	Phase
Marmousi onshore 3 Hz	α,β	λ,μ	κ,μ
Marmousi onshore 4 Hz	α,β	$\kappa,\mu*$	α,β
Marmousi onshore 5 Hz	α,β	α,β	—
Marmousi onshore 6 Hz	$\alpha,\beta*$	α,β	—
Marmousi onshore 7 Hz	κ,μ	α,β	—
overthrust onshore 3 Hz	α,β	ϕ,β	α,β
overthrust onshore 4 Hz	α,β	$\alpha,\beta*$	α,β
overthrust onshore 5 Hz	α,β	α,β	$\alpha,\beta*$
overthrust onshore 6 Hz	α,β	α,β	κ,μ
overthrust onshore 7 Hz	α,β	λ,μ	$\kappa,\mu*$
overthrust offshore 3 Hz	α,β	α,β	α,β
overthrust offshore 4 Hz	—	$\kappa,\mu*$	α,β
overthrust offshore 5 Hz	—	κ,μ	—
overthrust offshore 6 Hz	—	α,β	—
overthrust offshore 7 Hz	—	α,β	—
BP 2007 diapir 3 Hz	—	κ,μ	α,β
BP 2007 diapir 4 Hz	—	λ,μ	—
BP 2007 anticline 3 Hz	—	α,β	ϕ,β
BP 2007 anticline 4 Hz	—	α,β	α,β
BP 2007 anticline 5 Hz	—	$\alpha,\beta*$	—
Marmousi offshore 3 Hz	—	ϕ,β	λ,μ
Marmousi offshore 4 Hz	—	α,β	—

Table 4: Worst-performing material parameterization in terms of model error reduction (ΔE). Asterisk indicates a near tie. Dash denotes all parameterizations failed.

REFERENCES

Akçelik, A., G. Biros, and O. Ghattas, 2002, Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation.

Bozdag, E., J. Trampert, and J. Tromp, 2011, Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements: Geophysical Journal International, **185**, 845–870.

Bunks, C., M. Fatimetou, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Cai, X.-C., and D. E. Keyes, 2002, Nonlinearly preconditioned inexact Newton algorithms: SIAM Journal on Scientific Computing, **24**, 183–200.

Dennis, J., and R. Schnabel, 1996, Numerical methods for unconstrained optimization and nonlinear equations: SIAM.

Komatitsch, D., and J.-P. Vilotte, 1998, The spectral element method: an efficient tool to simulate the seismic response of 2D and 3D geologic structures: Bulletin of the Seismological Society of America, **88**, 368–392.

Liu, D., and J. Nocedal, 1989, On the limited memory BFGS method for large scale optimization: Mathematical Programming, **45**, 504–528.

Luo, Y., J. Tromp, B. Denel, and H. Calandra, 2013, 3d coupled acoustic-elastic migration with topography and bathymetry based on spectral-element and adjoint methods: Geophysics, **78**, S193–S202.

Métivier, L., F. Bretaudeau, R. Brossier, S. Operto, and J. Virieux, 2014, Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures: Geophysical Prospecting, 1–23.

Nash, S., and J. Nocedal, 1991, A numerical study of the limited memory BFGS method and the truncated-Newton method for large scale optimization: SIAM Journal of Optimization, **1**, 358–372.

Powell, M., 1977, Restart procedures for the conjugate gradient method: Mathematical Programming, **12**, 241–254.

Tape, C., Q. Liu, and J. Tromp, 2007, Finite-frequency tomography using adjoint methodsmethodology and examples using membrane surface waves: Geophysical Journal International, **168**, 1105–1129.

Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection data: Geophysics, **51**, 1893–1903.

Tromp, J., C. Tape, and Q. Liu, 2005, Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels: Geophysical Journal International, **160**, 195–216.

Yuan, Y. O., F. J. Simons, and E. Bozdag, 2015, Multiscale adjoint waveform tomography for surface and body waves: Geophysics, **80**, R281–R302.

Zhu, H., Y. Luo, T. Nissen-Meyer, C. Morency, and J. Tromp, 2009, Elastic imaging and time-lapse migration based on adjoint methods: Geophysics, **74**, WCA167–WCA177.