Data domain wavefield tomography using local correlation functions

Esteban Díaz and Paul Sava, Center for Wave Phenomena, Colorado School of Mines

SUMMARY

We formulate the data domain tomography misfit criterion using local correlations. Correlation-based inversions are less sensitive to local minima than difference-based inversions. Correlations, however, are often contaminated with cross-talk between events; in addition, the global correlations give just a general idea of the kinematic errors of the model because of the summation along the entire time axis. Alternatively, local correlations with Gaussian windows, advocated in this paper, are able to extract the local kinematic errors in the misfit between modeled and observed data. Local correlations are also less sensitive to cross-talk of seismic events than global correlations because the summation is performed locally as a function of time. Less correlation cross-talk leads to cleaner adjoint sources and hence, cleaner gradients. We further improve the gradients using a penalty function that is consistent with the bandwidth of the seismic data, which is more realistic than linear penalty functions designed to annihilate infinite bandwidth data.

INTRODUCTION

Wavefield tomography (WT) methods have the ability to recover velocity models that are consistent with the data bandwidth. This is because the model update is carried by wavepaths that are sensitive to the data bandwidth, as opposed to rays that are defined for high frequency asymptotic and that sample the model on an infinitesimal path. WT objective functions use the data difference as a measure of the similarity between observed and modeled data (Tarantola, 1984; Lailly, 1983; Pratt, 1999). This criterion is highly non-linear and often ill-posed, mainly bacause it compares oscillatory and band-limited functions; hence, an inversion can be trapped into a local minimum (Virieux and Operto, 2009).

Other waveform inversion methods sacrifice the resolution of the data difference for the robustness to errors beyond half a period. Luo and Schuster (1991) propose using the time-lag (measured at the maximum of the correlation between observed and modeled data) as the misfit criteria. van Leeuwen and Mulder (2010) demonstrate that penalizing the correlation, instead of picking its maximum, is more robust to phase difference errors due to the wrong choice of source functions to generate the modeled data (for which the maximum of the correlation does not match the true travel time error). Luo and Sava (2012) suggest to use deconvolution instead of correlation because the deconvolved misfit criterion offers better resolution than the conventional correlation approach. Warner and Guasch (2014) continue along the same line of deconvolution functionals through minimization of the non-zero lag components of the Wiener filter.

Perrone et al. (2015) use local correlation functions between migrated seismic images to formulate the wavefield tomography problem. Here, we propose to use local temporal correlation functions (Hale, 2006a) to extract the misfit information in the data domain (at the receiver locations) instead of measuring such correlations in a global sense, as suggested by other approaches. Local correlations minimize the cross-talk between seismic events, since the comparison between observed and simulated waveforms is performed locally in time. We demonstrate wavefield tomography implementation using the adjoint state method (Plessix, 2006) and illustrate the method with a synthetic example.

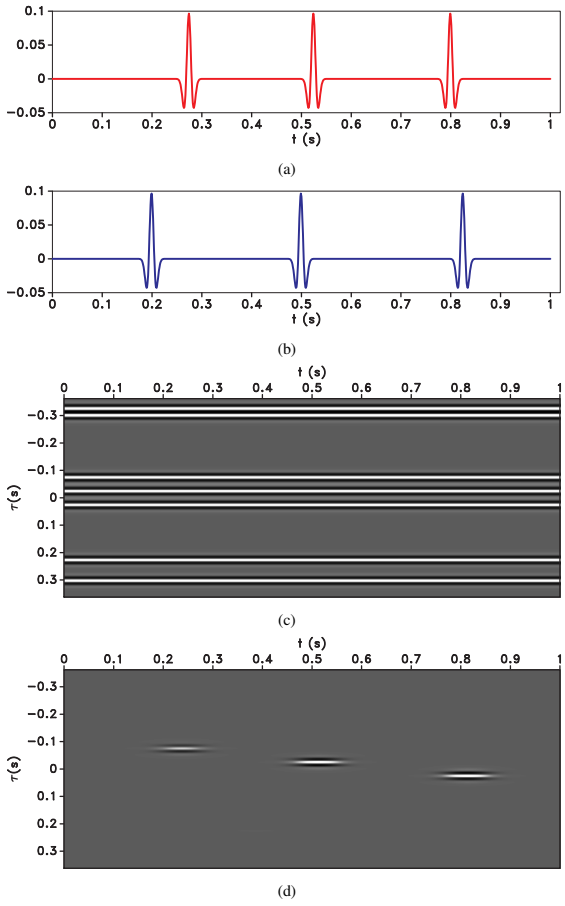

Figure 1: Signals (a) $f(t)$ and (b) $g(t)$ to exemplify the correlation functions; (c) is the global correlation (time invariant), and (d) is the local correlation with a Gaussian window with $\sigma = 0.3s$.

THEORY

Consider the problem of measuring the similarity between two signals $f(t)$ and $g(t)$, Figures 1a-1b. One way is to measure the difference between f and g. This approach is highly sensi-

tive if the two signals are close enough (in a kinematic sense) to each other. However, if the two signals contain a kinematic difference longer than half a period, this functional is not informative (Bunks et al., 1995). Instead of determining the similarity by measuring the difference, we can determine the similarity through correlation. Here, we follow the convention of Hale (2006a) and define centered local correlations as

$$c(t,\tau) = \int_{-\infty}^{\infty} w(t'-t) f(t'-\tau/2) g(t'+\tau/2) dt' \,, \qquad (1)$$

where $w(t'-t) \equiv e^{-\tau^2/4\sigma^2} e^{-(t'-t)^2/\sigma^2}$ is a Gaussian window centered at t', and σ is standard deviation of the Gaussian function. As $\sigma \to \infty$, the local correlations behave like global correlations at each time t, i.e. they are invariant with respect to time. This localized formulation is equivalent to convolving the shifted product $h_\tau(t') = f(t'-\tau/2)g(t'+\tau/2)$ with a Gaussian window: $c(t,\tau) = (G * h_\tau)(t)$. This convolution is efficiently implemented using recursive Gaussian filters (for which the convolution cost is independent of the Gaussian half-width σ) (Hale, 2006b). One can think about equation 1 as a linear operator $\mathbf{C} = \mathbf{GS}$ that applies the shifting operator $\mathbf{S}$ followed by the Gaussian convolutional operator $\mathbf{G}$. The adjoint operator $\mathbf{C}^\top$ implements a local convolution as $\mathbf{C}^\top = \mathbf{S}^\top \mathbf{G}$. By construction, the matrix representing operator $\mathbf{G}$ is SPD since all of the blocks are SPD: hence, $\mathbf{G}^\top = \mathbf{G}$.

Figure 1c shows the global correlation between the signals in Figures 1a and 1b. Note that the global correlation is time-invariant and contains cross-talk between events (around $\tau = -0.3$s and $\tau = 0.3$s); the cross-talk is captured because the correlation lags are big enough to facilitate interference between different events. In contrast, Figure 1d shows the local correlation between $f(t)$ and $g(t)$; no cross-talk is observed because the Gaussian window prevents interference between distant events.

Tomography formulation

Consider the acoustic wave equation

$$\frac{1}{v^2}\ddot{u} - \nabla^2 u = s \,, \qquad (2)$$

where $u(\mathbf{x},t)$ is the wavefield excited by the source function $s(\mathbf{x},t)$, and $v(\mathbf{x})$ is the medium P wave velocity. We can extract data from the wavefield by using a restriction operator at the receiver positions $\mathbf{x}_r$ as $d(\mathbf{x}_r,t) = u(\mathbf{x},t)\,\delta(\mathbf{x}-\mathbf{x}_r)$. For the remainder of this paper, we express the wave equation as the linear operator $\mathbf{L}(m)$ with its corresponding adjoint operator $\mathbf{L}^\top(m)$, where $m = 1/v^2$ is the model parameter representing slowness squared. We can design a correlation-based objective function that maximizes the correlation between observed and modeled data:

$$J(m) = \frac{||Pd^o \star_l d^m||_2^2}{||d^o \star_l d^m||_2^2} = \frac{||\mathbf{PCd^m}||_2^2}{||\mathbf{Cd^m}||_2^2}, \qquad (3)$$

where $P(\tau)$ is a penalty operator that enhances energy outside $\tau = 0$, m is the model parameter, and $d^o(e,\mathbf{x}_r,t)$ and $d^m(e,\mathbf{x}_r,t)$ are the observed and modeled data for each experiment e, respectively. The correlation matrix $\mathbf{C}$ is built using shifted ver-

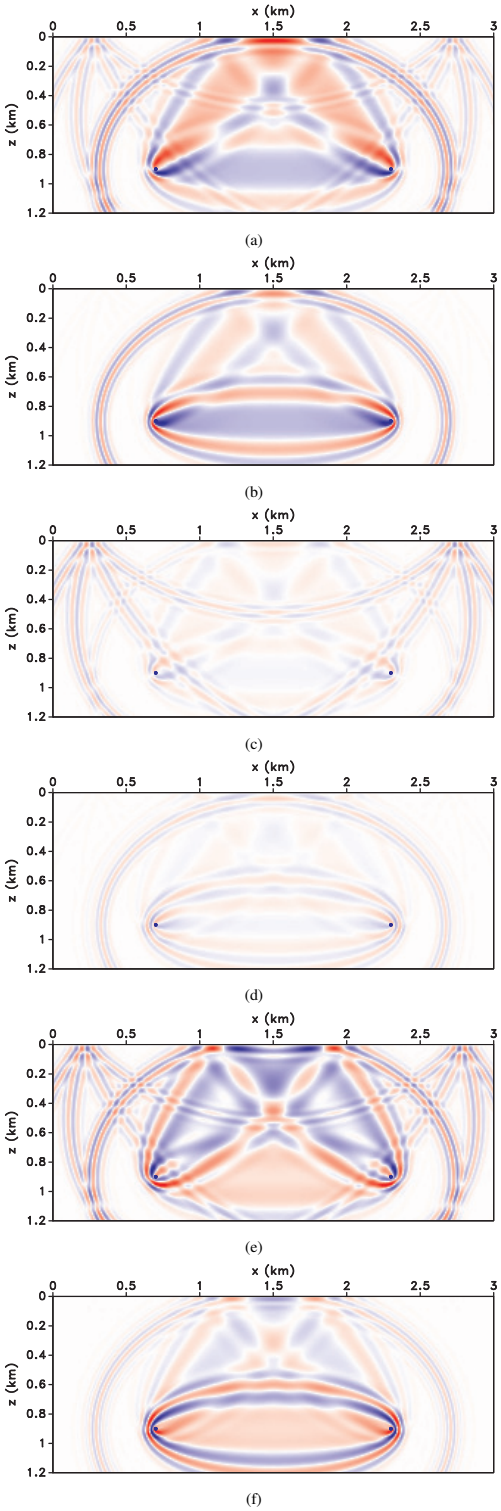

Figure 2: Sensitivity kernels for an experiment with free surface boundary condition and a Gaussian anomaly placed in between the source and the receiver: (a) global and (b) local correlation gradients for a negative Gaussian anomaly, (c) global and (d) local correlation gradients for the correct model, and (e) global and (f) local correlation gradients for a positive Gaussian anomaly.

sions of the observed data d^o along with the Gaussian convolution, as defined by the process in equation 1. The application of the penalty operator $P(\tau)$ is equivalent to a diagonal penalty matrix $\mathbf{P}$. Later, we discuss strategies for choosing the penalty operator $P(\tau)$.

We solve the optimization problem using gradient descent methods. The adjoint state method framework (Plessix, 2006) allows us to obtain the gradient of our constrained problem as

$$H(u_s, a_s, m) = J(m) - \sum_e \langle a_s, Lu_s - s \rangle \ , \tag{4}$$

from which we seek to obtain the appropriate Lagrange multipliers $a_s(e,\mathbf{x},t)$ (adjoint wavefields) that satisfy the constrained equation. We use the notation $\langle .,. \rangle$ for the dot product between two vectors.

The stationary points of the Lagrangian provide us with the equations to solve the problem. The first condition $\partial_{a_s} H = 0$ leads to the state variable problem

$$\mathbf{L}(m) u_s(e,\mathbf{x},t) = s(e,\mathbf{x},t) \tag{5}$$

for each experiment e. The second condition $\partial_{u_s} H = 0$ leads to the adjoint state variable system

$$\mathbf{L}^\top(m) a_s(e,\mathbf{x},t) = g_s(e,\mathbf{x},t) \ , \tag{6}$$

where $g_s(e,\mathbf{x},t)$ is the adjoint source for our specific choice of objective function. For data comparison through correlation, the adjoint sources are given by

$$g_s = -\frac{1}{||\mathbf{C}d^m||_2^2} \mathbf{C}^\top \left(\mathbf{P}^\top \mathbf{P} - J\mathbf{I} \right) \mathbf{C} d^m \ . \tag{7}$$

Once we solve for u_s and a_s, we can compute the gradient of J with the last condition $\partial_m H = 0$, from which we obtain the gradient

$$\partial_m J = \sum_{e,t} \ddot{u}_s a_s \ . \tag{8}$$

Penalty operators

The penalty operator $P(\tau)$ has the objective of enhancing correlation energy away from zero time-lag. A conventional penalty function is $P(\tau) = |\tau|$ (Shen and Symes, 2008; van Leeuwen and Mulder, 2010; Luo and Sava, 2012) based on the concept of differential semblance optimization (Symes, 2008), and is designed to annihilate energy linearly as a function of τ. Alternatively, we can construct a penalty function that is consistent with the bandwidth of the data, similar to the procedure used by Yang et al. (2013) for image-domain wavefield tomography. Thus, we define the penalty using the auto-correlation of the observed data:

$$P(\tau) = \sum_{\mathbf{x}_r} \frac{|\tau|}{Env(d^o \star d^o(\tau, \mathbf{x}_r)) + \varepsilon} . \tag{9}$$

The envelope function Env captures the energy of the auto-correlation, and ε is a stabilization factor for the division.

We illustrate the method with a simple model with constant velocity and with free surface. We perturb the medium with a Gaussian anomaly in such a way that the free surface reflection arrival time is not affected. Figures 2a, 2c, and 2e show the gradients from the global correlations. Strong cross-talk between events is visible as high wavenumber (reflection-like) events coincident at the free surface and elsewhere. However, the gradients from local correlations (Figures 2b, 2d, and 2f) show less contamination from cross-talk.

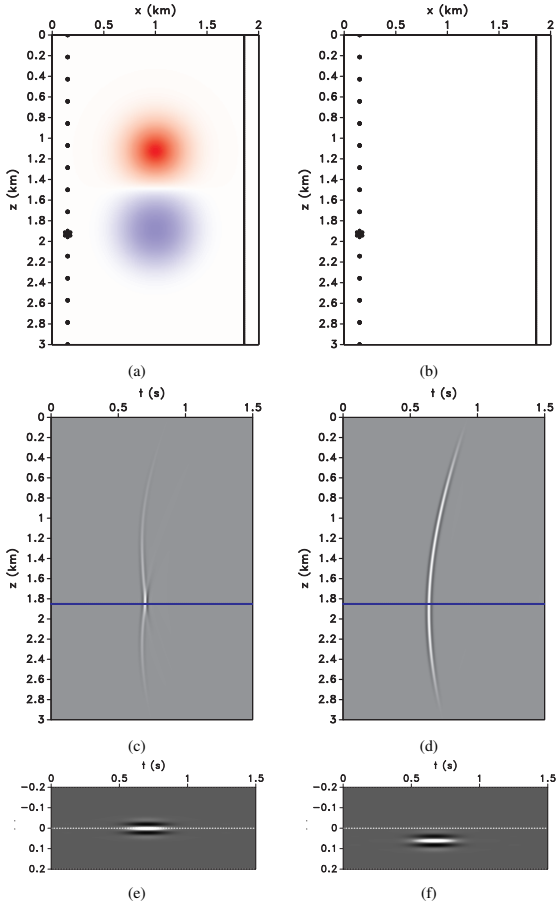

Figure 3: Wavefield tomography example with two Gaussian anomalies of opposite sign: (a) the true model, (b) the starting model (constant background), (c) the observed data from the highlighted shot point at $z = 1.9$km, (d) the modeled data from the background model, (e) the local auto-correlation of the observed data, and (f) the local correlation of initial data with observed data. The horizontal line indicates the location of the correlation functions.

EXAMPLE

We illustrate our method with a transmission example, as shown in Figure 3a. The model contains a background velocity of 3km/s and two Gaussian anomalies of $\pm$0.6km/s. The dots on the left of the model show the source locations, and the line on the right of the model shows the receiver line. We perform inversion with a starting model (Figure 3b) equal to the constant background velocity of 3km/s. The two Gaussian anomalies

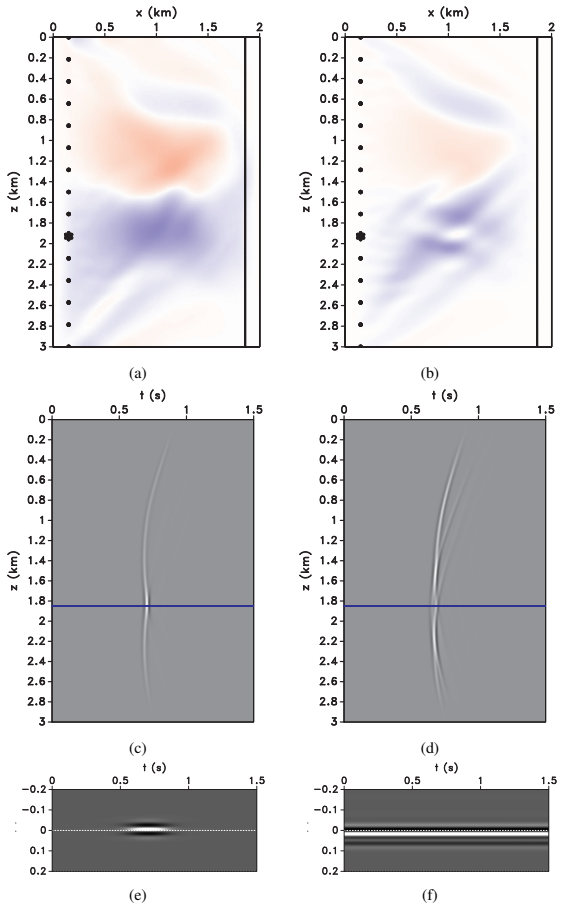

Figure 4: Inversion results for the example in Figure 3: (a) the inverted model from local correlations, (b) the inverted model from global correlations, (c) the simulated data from model (a), (d) the simulaated data from the model (b), (e) the correlation at the final iteration for local correlation inversion, and (f) the correlation at the final iteration for the global correlation inversion. The horizontal line indicates the location of the correlation functions.

produce triplications in the wavefield as shown by the observed data in Figure 3c. The data from the starting model is shown in Figure 3d. Figures 3e-3f show the auto-correlation of observed data for a particular receiver at $z = 1.85$km, and the correlation of the initial data with observed data, respectively.

Figure 4a shows the retrieved model from using the local correlation objective function, whereas Figure 4b shows the less well resolved model from the global correlation. Figures 3c, 4c, and 4d show the modeled data for the highlighted source point in the true model, local correlation model, and global correlation models, respectively. Even though the global correlation inversion recovers the correct sign of the anomaly, it is not able to create data similar to that of the true model. In contrast, the data from the local correlation model correctly recovers the triplication produced by the low velocity zone located in the lower half of the model. Figures 4e-4f show the correlation functions for the model retrieved with local correlations, and the model retrieved with global correlations, respectively. The correlation in Figure 4e obtained from the model in Figure 4a matches the auto-correlation in Figure 3e. Figure 5 shows the convergence curves for the local correlation inversion (solid line) and the global correlation inversion (dashed line). The local correlation inversion converges significantly faster than the global correlation inversion.

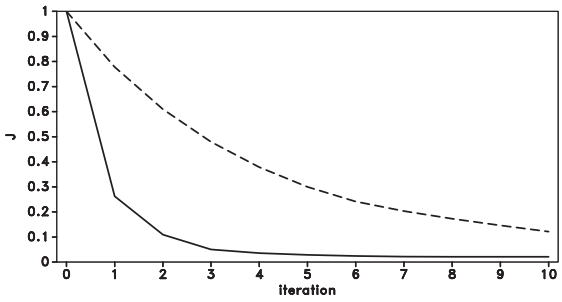

Figure 5: Convergence curves for the Gaussian anomalies model for local correlations (solid line) and global correlations (dashed line). Note how the local correlation convergence is faster than the global correlation inversion.

CONCLUSIONS

Local as opposed to global correlations provide cleaner measures of similarity between two functions. The main difference is that local correlation performs a local sum along the time axis. This is the key to eliminating the comparison between unrelated seismic events present in the observed and modeled data. This feature allows us to extract local kinematic errors that are more instructive than the misfit extracted using global correlations. Less cross-talk in the correlations translates into cleaner adjoint sources, which can produce more informative model updates than those based on global correlation.

The proposed penalty function ensures faster convergence since it annihilates events in the correlation function that are consistent with the data bandwidth. Once the local correlation misfit is minimum, one can continue with high resolution functionals based on the data difference, i.e., waveform inversion. We also show that local correlations together with bandlimited penalty functions produce cleaner gradients and converge faster than a comparable inversion using global correlations. Our adjoint operator for the local correlations method suggests several possible applications including objective functions based on local deconvolution, which extends further to local-matched filters applications.

ACKNOWLEDGMENTS

We thank the sponsor companies of the Consortium Project on Seismic Inverse Methods for Complex Structures. The reproducible numeric examples in this paper use the Madagascar software package (Fomel et al., 2013).

REFERENCES

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

Fomel, S., P. Sava, I. Vlad, Y. Liu, and V. Bashkardin, 2013, Madagascar: open-source software project for multidimensional data analysis and reproducible computational experiments: Journal of Open Research Software, **1**, e8.

Hale, D., 2006a, An efficient method for computing local cross-correlations of multi-dimensional signals: CWP, **544**, 1–7.

———, 2006b, Recursive gaussian filters: CWP-546.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on inverse scattering: theory and application, Society for Industrial and Applied Mathematics, Philadelphia, PA, 206–220.

Luo, S., and P. Sava, 2012, A deconvolution–based objective function for wave–equation inversion: SEG Technical Program Expanded Abstracts 2011, 2788–2792.

Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**, 645–653.

Perrone, F., P. Sava, and J. Panizzardi, 2015, Waveform tomography based on local image correlations: Geophysical Prospecting, **63**, 35–54.

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**, 495–503.

Pratt, R., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model: Geophysics, **64**, 888–901.

Shen, P., and W. W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, VE49–VE59.

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.

van Leeuwen, T., and W. A. Mulder, 2010, A correlation-based misfit criterion for wave-equation traveltime tomography: Geophysical Journal International, **182**, 1383–1394.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.

Warner, M., and L. Guasch, 2014, Adaptive waveform inversion: Theory: SEG Technical Program Expanded Abstracts 2014, 1089–1093.

Yang, T., J. Shragge, and P. Sava, 2013, Illumination compensation for image-domain wavefield tomography: Geophysics, **78**, U65–U76.