Seismic tomography using local correlation functions

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

ABSTRACT

Wavefield tomography in the data domain is usually formulated by using the data difference as the misfit criterion. The data difference, however, is susceptible to cycle skipping which leads to convergence into local minima. To overcome the strong non-linearity, a multi-scale approach to the inversion is often needed. The success of this approach relies on the low-frequency content of the data. Here, we define 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 their global correlations counterpart 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.

Key words: Wavefield tomography, local correlations, inversion

1 INTRODUCTION

Waveform tomography (WT) methods have the ability to recover velocity models that are consistent with the data bandwidth. Kernels of wavefield tomography contain higher sensitivity to low wavenumbers compared to those of rayfield tomography. This is because the update is carried by wavepaths that are sensitive to the bandwidth of the data, as opposed to the update carried by rays that are defined in the context of high frequency asymptotic and that sample the model on an infinitesimal path. Typically, objective functions of WT 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 ill-posed (Virieux and Operto, 2009), mainly due to the fact that it compares oscillatory and band-limited functions; hence, an inversion can quickly fall into a local minima if the traveltime error is smaller than half a period (Virieux and Operto, 2009). This non-linearity requires a multi-scale approach where the model is updated from low wavenumbers to high wavenumbers (Bunks et al., 1995), and this approach often imposes serious low-frequency requirements on the data used for FWI.

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 to use 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 (correlation with the inverse of the modeled data) as opposed to direct correlation and assert that the deconvolved misfit criterion contains 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, which are computed from the observed data to match the modeled data. All these waveform tomography methods are optimized with gradient descent minimization methods. The gradients are computed efficiently subject to wave equation constraints implemented within the framework of the adjoint state method (Plessix, 2006).

Perrone et al. (2015) use local correlation functions be-

tween migrated seismic images to define the tomography problem. Here, we propose to use local temporal correlation functions 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. To compute the local correlation functions, we use the method of Hale (2006a), which performs the local summation in an efficient manner using recursive Gaussian filters (Hale, 2006b). Local correlations minimize the cross-talk between seismic events, since the comparison between observed and simulated waveforms is performed locally in time. This is a key advantage of our method because it has the ability to produce cleaner adjoint sources, which translate into cleaner gradients (with less cross-talk than the global correlation functionals).

This paper is organized as follows: we start by reviewing the theory behind global and local-correlation functions, where we also define the corresponding adjoint pair for the local-correlation linear operator. Then, we introduce the theory for the wavefield tomography problem based on our new functional, and we show its corresponding implementation using the adjoint state method (Plessix, 2006). We then illustrate the method with a synthetic example and discuss different approaches for applying penalization functions to global and local correlations.

2 CORRELATION FUNCTIONALS.

Consider the problem of measuring the similarity between two signals $f(t)$ and $g(t)$ (Figures 1(a)-1(b)). One way is to measure the difference between f and g. This approach is highly sensitive 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. In the most common case, one can use a global correlation function (van Leeuwen and Mulder, 2010). Here, we follow the convention of Hale (2006a) and define centered correlations:

$$c(\tau) = (f \star g)(\tau) = \int_{-\infty}^{\infty} f(t' - \tau/2) g(t' + \tau/2) dt'. \quad (1)$$

The correlation function $c(\tau)$ gives us a general sense of the similarity between $f(t)$ and $g(t)$. Alternatively, we can estimate the similarity in a localized way (Hale, 2006a), where we can capture how the two signals are related to each other along the time axis:

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

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 the global correlations at each time t, i.e. they are invariant with respect to time t. 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: $c(t,\tau) = (G * h_\tau)(t)$. The 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 2 as a linear operator defined as $\mathbf{c} = \mathbf{C}\mathbf{g} = \mathbf{G}\mathbf{S}\mathbf{g}$, where $\mathbf{C} = \mathbf{G}\mathbf{S}$ is the local correlation matrix, $\mathbf{G}$ is a block diagonal matrix (with each block being a convolution matrix with the Gaussian function), and $\mathbf{S}$ contains shifted versions of f. The transposed (adjoint) operator $\mathbf{C}^\top$ implements a local convolution as $\mathbf{C}^\top = \mathbf{S}^\top \mathbf{G}$. By construction, the matrix $\mathbf{G}$ is SPD since all of the blocks are SPD: hence, $\mathbf{G}^\top = \mathbf{G}$. Mathematically, the adjoint operator is implemented as follows:

$$f^*(t) = \int_{\tau_{min}}^{\tau_{max}} \left(\int_{-\infty}^{\infty} w(t' - t + \tau/2) c(t' + \tau/2, \tau) dt' \right) g(t+\tau) d\tau, \quad (3)$$

where τ_{min} and τ_{max} are the minimum and maximum time-lags, respectively, and f^* represents the adjoint of f.

Figure 1(c) shows the global correlation between the signals in Figures 1(a) and 1(b). 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 1(d) shows the local correlation between $f(t)$ and $g(t)$; no cross-talk is observed because the Gaussian window prevents interference between distant events. However, it is still possible to observe cross-talk between events using local correlations if the events lie within the correlation window of standard deviation σ.

3 TOMOGRAPHY WITH CORRELATION FUNCTIONS

Here, we show the application of correlation functionals for the inverse tomographic problem under the physics of the acoustic wave equation

$$\frac{1}{v^2} \frac{\partial^2 u}{\partial t^2} - \nabla^2 u = s, \quad (4)$$

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 $m = 1/v^2(\mathbf{x})$ is the model parameter representing the slowness squared. We can design a correlation-based objective function that maximizes the correlation between observed and modeled data:

$$J(m) = \frac{||P(\tau) d^o \star_l d^m||_2^2}{||d^o \star_l d^m||_2^2} = \frac{||\mathbf{P}\mathbf{C}\mathbf{d}^\mathbf{m}||_2^2}{||\mathbf{C}\mathbf{d}^\mathbf{m}||_2^2}, \quad (5)$$

where $P(\tau)$ is a penalty operator that enhances energy outside $\tau = 0$ (which we want to minimize), 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

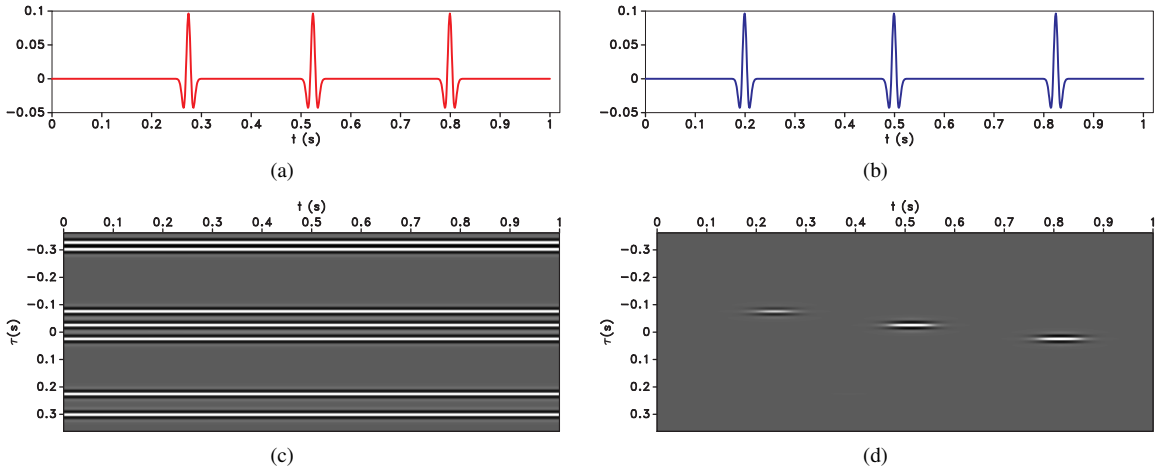


Figure 1. Signals (a) $f(t)$ and (b) $g(t)$ to exemplify the correlation functions, (c) is the global correlation $f \star g$ (invariant with time), and (d) $f \star_l g$ is the local correlation with a Gaussian window with $\sigma = 0.3s$.

modeled data for each experiment e, respectively. The correlation matrix $\mathbf{C}$ is built using shifted versions of the observed data d^o along with the Gaussian convolution, as defined by the process in equation 2. The application of the penalty operator $P(\tau)$ is equivalent to a diagonal penalty matrix $\mathbf{P}$. Here, the L2 norm squared $||.||_2^2$ is defined over receiver coordinates $\mathbf{x}_r$, experiment index e, time t, and correlation lag τ. Later, we discuss strategies for choosing the penalty operator $P(\tau)$.

The objective function in equation 5 is similar to that of van Leeuwen and Mulder (2010) except that we use local instead of global correlation, and we normalize the denominator. The normalization balances the objective function by the energy of the correlation function (Warner and Guasch, 2014). This step eliminates the objective function bias for positive and negative errors.

To solve the optimization problem, we use gradient descent methods. The adjoint state method framework (Plessix, 2006) allows us to obtain the gradient by defining our problem under equality constraints given by the wave equation as

$$\begin{aligned} \underset{m}{\text{minimize}} \quad & J(m) \\ \text{subject to} \quad & L_e u = s. \end{aligned} \tag{6}$$

These constraints satisfy the wave equation for the modeled data d^m for every experiment e. We define the augmented functional as

$$H(u_s, a_s, m) = J(m) - \sum_e < a_s, Lu_s - s >, \tag{7}$$

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 $< ., . >$ 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

$$\partial_{a_s} H = 0 \tag{8}$$

leads to the state variable problem

$$\mathbf{L}(m) u_s(e, \mathbf{x}, t) = s(e, \mathbf{x}, t) \tag{9}$$

for each experiment e. The second condition

$$\partial_{u_s} H = 0 \tag{10}$$

leads to the adjoint state variable system

$$\mathbf{L}^\top(m) a_s(e, \mathbf{x}, t) = g_s(e, \mathbf{x}, t), \tag{11}$$

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

$$g_s = -\frac{1}{||\mathbf{C}d^m||_2^2} \mathbf{C}^\top \left(\mathbf{P}^\top \mathbf{P} - J\mathbf{I} \right) \mathbf{C}d^m. \tag{12}$$

This expression tells us to twice penalize the correlation and subtract the value of the objective function from it before computing the adjoint correlation. This process effectively balances the influence between the numerator and denominator in the adjoint source. At early iterations, when J is larger, the norm of the correlation has greater influence on the adjoint source; as iterations continue, the penalty term becomes more prominent. The correct calculation of the adjoint source relies on the fact that the linear operators and their corresponding transpose (adjoint) operators are correctly implemented. We make sure this is the case by applying the dot product test (Claerbout, 2008) on each linear operator. Once we solve for u_s and a_s, we can compute the gradient of J with the last condition

$$\partial_m H = 0, \tag{13}$$

from which we obtain the gradient

$$\partial_m J = \sum_{t,e} \ddot{u}_s a_s, \tag{14}$$

where the double dot indicates the second derivative in time.

3.1 Penalty operators

The penalty operator $P(\tau)$ has the objective of enhancing the energy that does not comply with our assumption of maximum correlation near zero lag for a correct velocity model. A conventional penalty function is $P(\tau) = |\tau|$ (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. This penalty function goes to zero only if the data, and thus the correlation, has infinite bandwidth. This linear weight becomes less sensitive as the data bandwidth decreases. For band-limited data, this penalty can produce high residuals even for a correct model. 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:

$$P(\tau) = \sum_{\mathbf{x}_r} \frac{1}{Env(d^o \star d^o(\tau, \mathbf{x}_r)) + \epsilon}. \qquad (15)$$

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 a 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 2(a), 2(c), and 2(e) show the models for positive, zero, and negative Gaussian anomaly, respectively. Figures 2(b), 2(d), and 2(f) show the corresponding data, for positive, zero, and negative Gaussian anomalies. Figures 3(a), 3(c), and 3(e) show the global correlations for negative, correct, and positive perturbations, respectively. One can see the cross-talk in the correlations for lags longer than $\tau = \pm 0.1$s. Figures 3(b), 3(d), and 3(f) depict the local correlations for the same events. In this case, the cross-talk is greatly attenuated, and the correlation events are localized without interference along the time axis. The cross-talk in the global correlations is propagated into the corresponding gradients through the adjoint sources. Figures 4(a), 4(c), and 4(e) 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 4(b), 4(d), and 4(f)) are cleaner and present the correct gradient direction without the contamination from cross-talk, which is almost non-existent for the chosen window. In this simple setting, we expect two sensitivity paths: (1) from source to receiver, and (2) from source to the free surface and from the free surface to the receivers. However, we can observe several more events in the global correlation gradients. These extra events come from the cross-talk present in global correlations, which is then propagated into the adjoint wavefields.

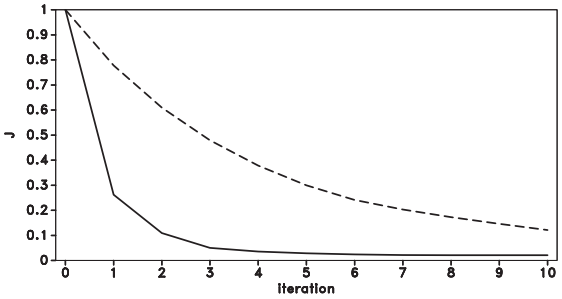


Figure 7. Convergence curves for the Gaussian anomalies model for local correlations (solid line) and global correlations (dashed line). Note how the local correlation convergence is faster than the global correlation inversion.

4 TRANSMISSION TOMOGRAPHY WITH GAUSSIAN ANOMALIES

In this example, we use a transmission setting with the model shown in Figure 5(a). The model contains a background velocity of 3km/s and two Gaussian anomalies of ± 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 5(b)) equal to the constant background velocity of 3km/s. The two Gaussian anomalies produce triplications in the wavefield as shown by the observed data in Figure 5(c). The data from the starting model is shown in Figure 5(d). Figures 5(e)-5(f) 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 6(a) shows the retrieved model from using the local correlation objective function, whereas Figure 6(b) shows the inverted model from the global correlation. One can observe that the model is better retrieved with the local correlation framework. Figures 5(c), 9(e), and 9(f) show the modeled data for the highlighted source point in the true model, local correlation model, and global correlation model, respectively. Even though the global correlation inversion recovered 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 6(e)-6(f) show the correlation functions for the model retrieved with local correlations, and the model retrieved with global correlations, respectively. We can see that the retrieved model with global correlations produces extra events when compared to the correlation from the local correlation inversion, which is simpler. The correlation in Figure 6(e) obtained from the model in Figure 6(a) matches the auto-correlation in Figure 5(e). Figure 7 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.

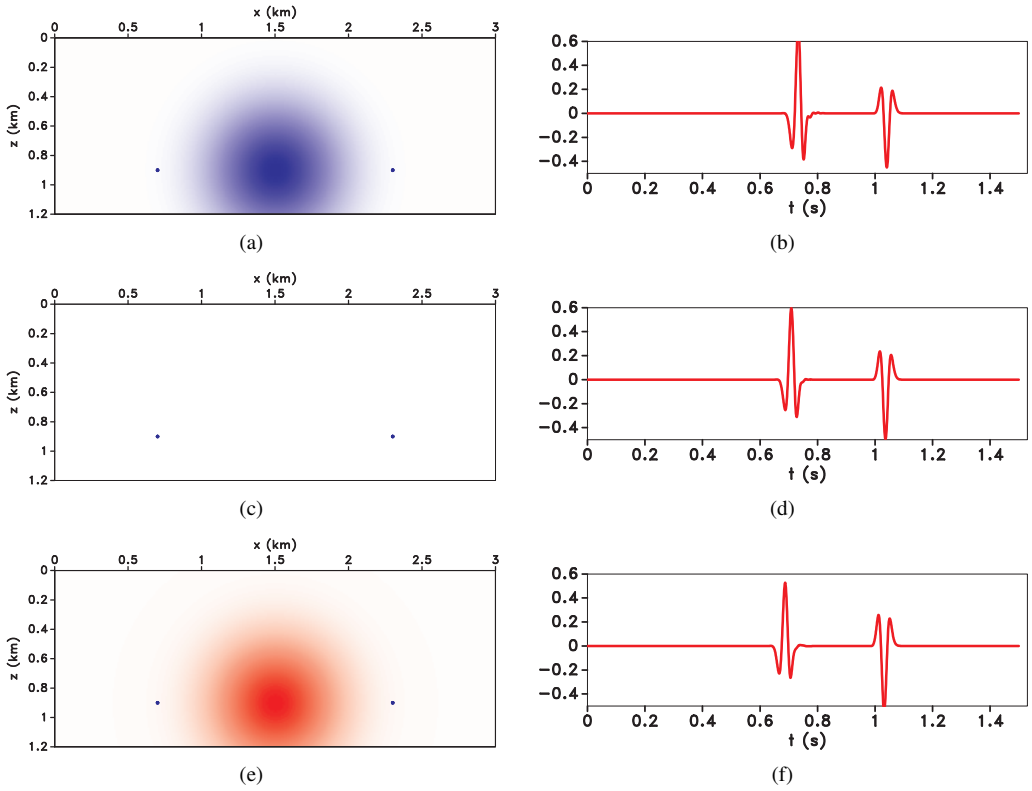


Figure 2. Sensitivity kernel experiment: the left column shows the models for (a) negative, (c) zero, and (e) positive Gaussian anomaly. The right column (b), (d), and (f) show the corresponding modeled data. For computing both the kernels we use the constant background model.

5 TRANSMISSION TOMOGRAPHY WITH THE MARMOUSI MODEL

We also demonstrate our method with a portion of the Marmousi–2 model (Martin et al., 2002) in a transmission setup. We slightly smooth the original model to create the true model used in this example (Figure 8(a)). The acquisition geometry consists of 14 shots located at $x = 8.12$km equally spaced in the vertical direction. On the right hand side of the model, we place the receiver line at $x = 12.5$km. Figure 8(c) shows the data for the shot at $\mathbf{x}_s = (8.12, 1.95)$km.

We strongly smooth the true model to create the starting model, Figure 8(b). Figures 8(c) and Figure 8(d) show the observed and modeled data from the true and starting models, respectively. One can see that the initial model is only able to generate a simple arrival, which differs from the the observed data shown in Figure 8(c). Figures 8(e)-8(f) show the correlation functions for the data corresponding to the true model (which is an auto-correlation) and the starting model, correspondingly.

Figures 9(a)-9(b) show the inverted models for local and global correlations, respectively, and Figures 9(c)-9(d) show the gradients at the first iteration, correspondingly; One can observe that the gradient from the local framework contains smoother features, whereas the global correlation gradient contains higher wavenumber events which arise from cross-talk injected in adjoint sources. Figures 9(e)-9(f) show the corresponding modeled data for local and global correlations, respectively. One can observe that the inverted models adds complexity into the seismic arrivals, which correctly matches the phase of the early arrivals from the observed data (Figure 8(c)). However, both models are unable to retrieve most of the later arrivals from the observed data. One reason for this behavior could be the strong gradient in the velocity model, which can produce poor illumination in the shallow section of the model. In the deeper part of the model, both inversions (in particular, the one from the local correlation framework) recover the high velocity layer near the reservoir. Underneath the reservoir, both inversions introduce a low velocity zone, which is also present in the true velocity model. Figures 9(g)-9(h) show the correlation functions in the inverted models, for local and global correlations, correspondingly. We can see how both of the inverted models produce correlations closer to $\tau = 0$ compared to those of the starting model. Figure 10 shows the convergence curves for the local correlation (solid line) and for the global correlation inversion (dashed line). One can observe (again) the improved convergence from the local correlation framework.

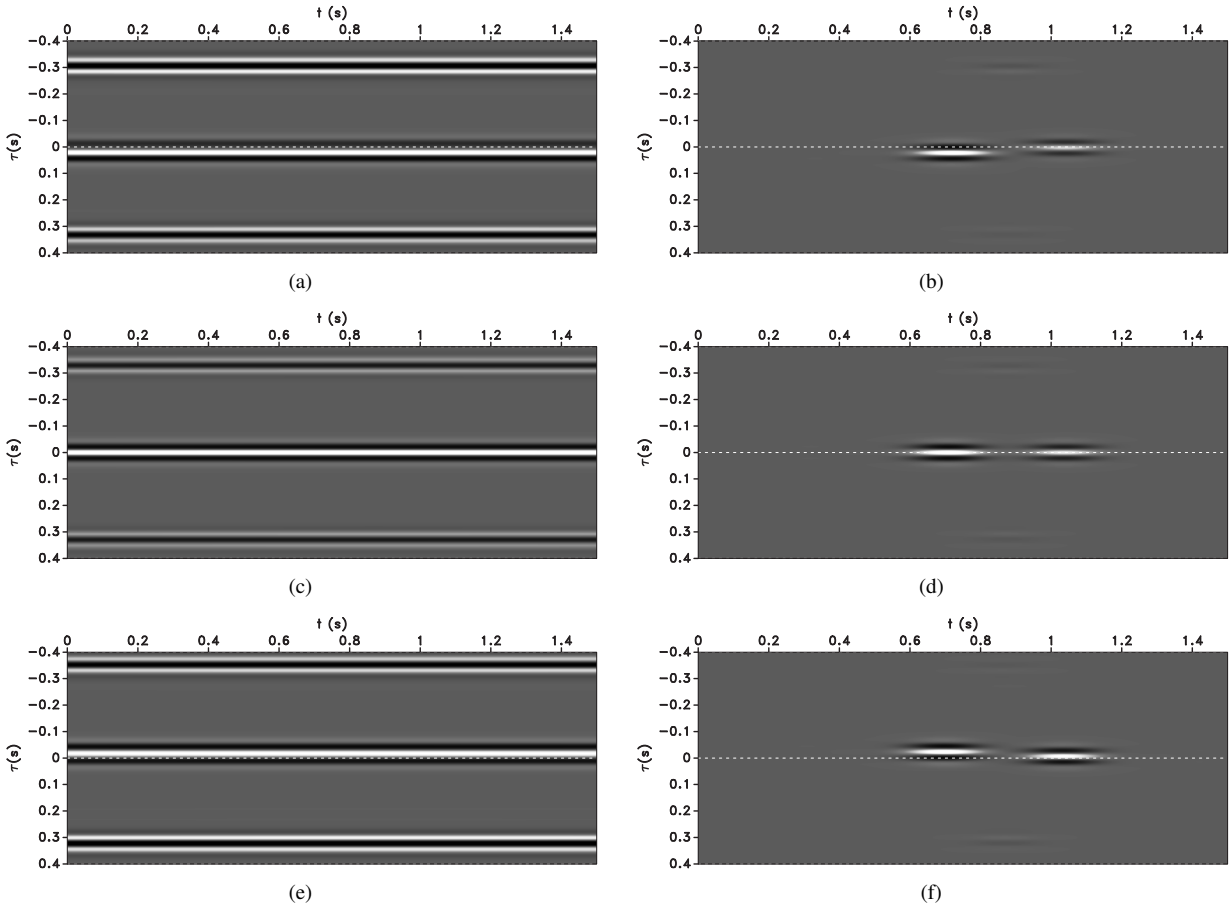


Figure 3. Sensitivity kernel experiment: the left and right columns show global and local correlations, respectively. From top to bottom one can see correlations for negative, zero, and positive Gaussian perturbations, respectively. Note how the local correlations clearly show the kinematic errors for the direct arrival event, whereas the global correlation shows the two events together. Also, the global correlation contains strong cross-talk between the two different events.

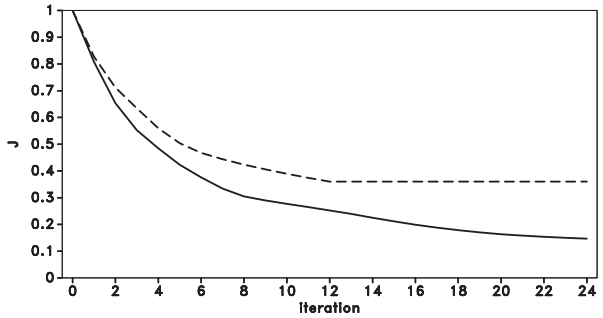


Figure 10. Convergence curves for the Marmousi model for local correlations (solid line) and global correlations (dashed line). Note how the local correlation convergence is faster than the global correlation inversion.

6 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 instead of summing over all the time samples. 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 of global correlation framework.

We also propose to use a bandlimited penalty function, which is constructed from the auto-correlation of the source function. This 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 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 sug-

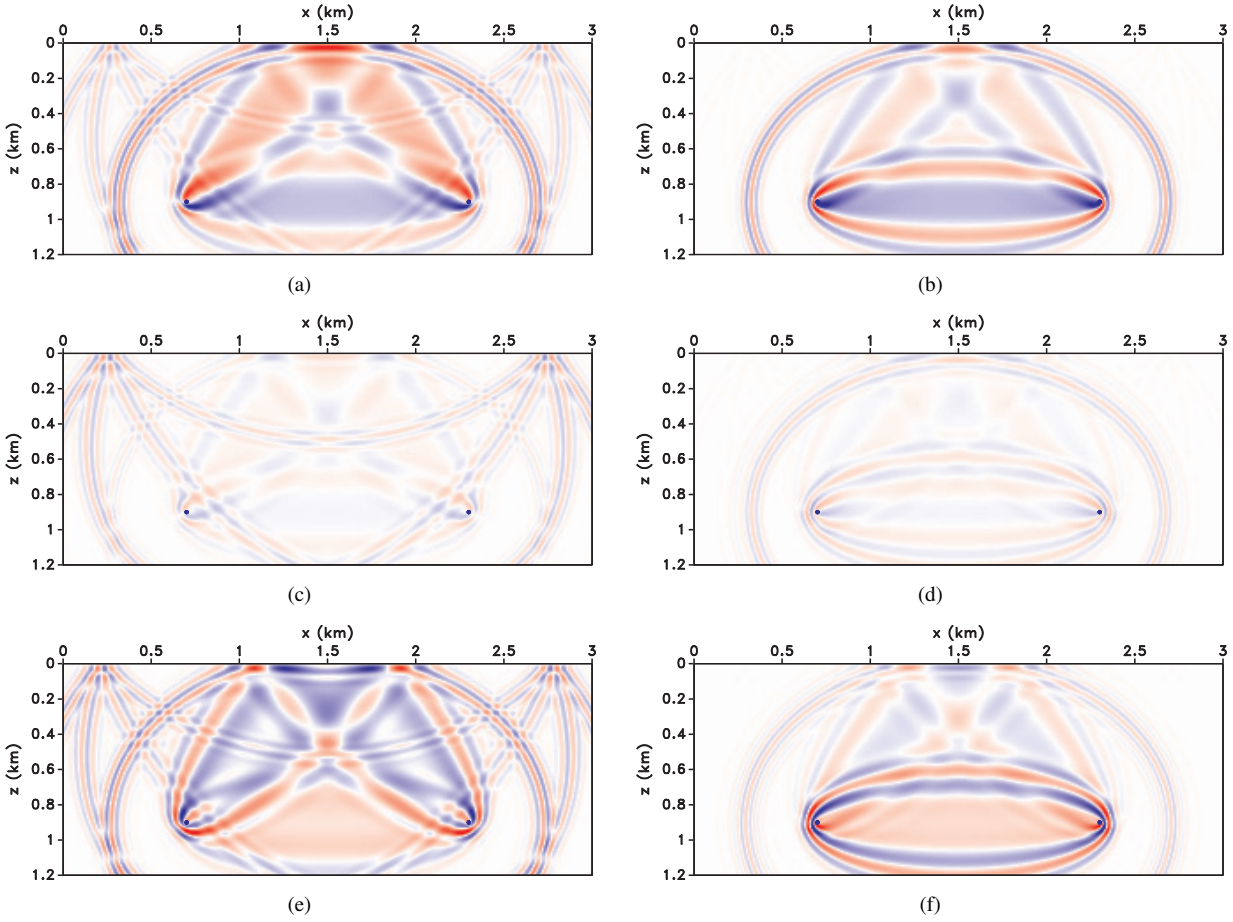


Figure 4. Sensitivity kernels for an experiment with free surface boundary condition and a Gaussian anomaly placed in between the source and the receiver (for which the surface reflection event is not sensitive). The left and right columns show gradients for global and local correlations, respectively. From top to bottom, the gradients correspond to low, correct, and high velocity.

gests several possible applications including objective functions based on local deconvolution, which extends further to local-matched filters applications.

7 ACKNOWLEDGMENTS

We thank 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.

Claerbout, J. F., 2008, *in* Basic earth imaging: Stanford University, 11–22.

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.

Martin, G. S., K. J. Marfurt, and S. Larsen, 2002, Marmousi–2: an updated model for the investigation of avo in structurally complex areas: SEG Expanded Abstracts, 1979–

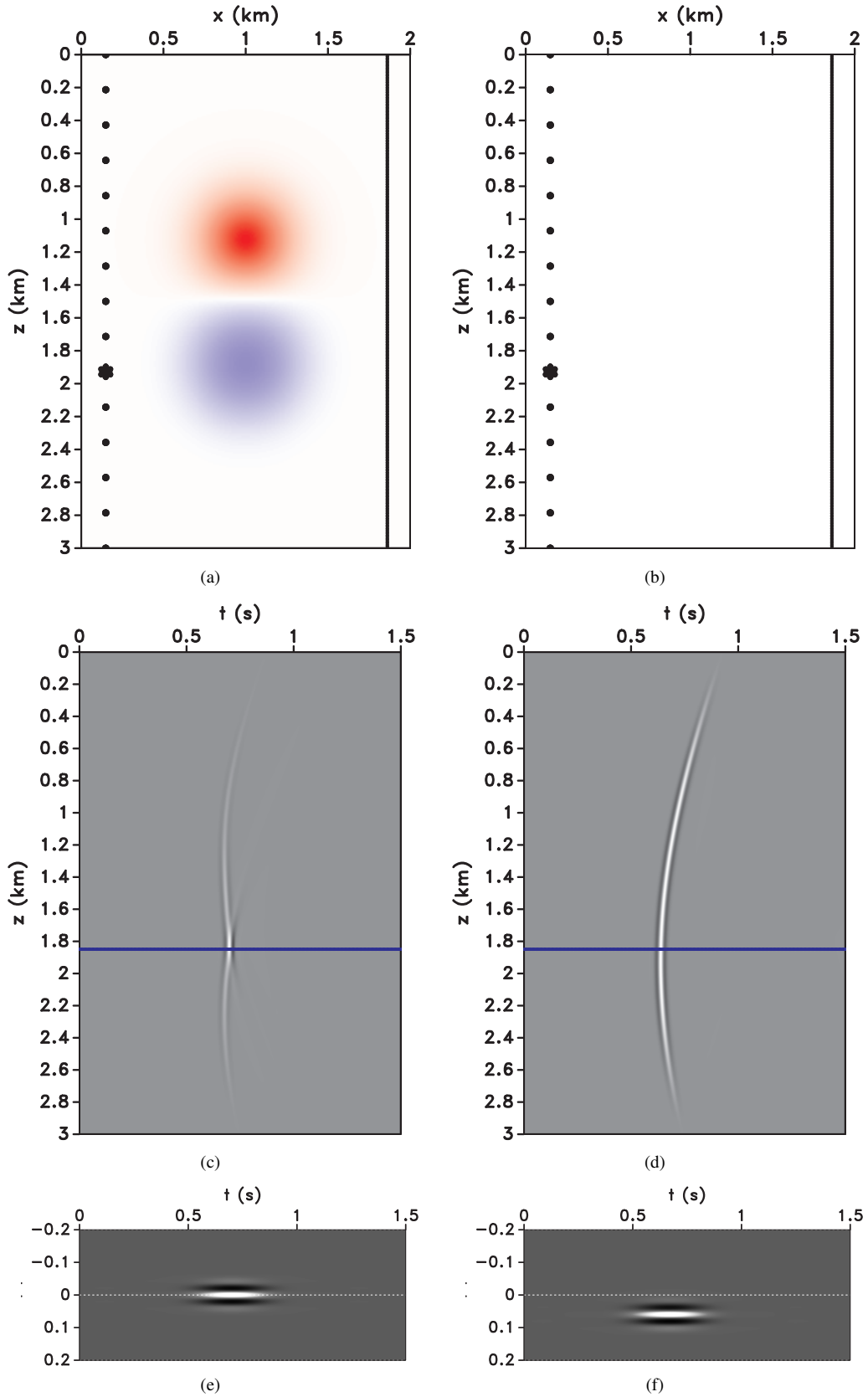


Figure 5. Transmission example with two Gaussian anomalies of opposite sign setup: (a) shows the true model, (b) the starting model (constant background velocity), (c) depicts the observed data from the highlighted shot point at $z = 1.9$km, (d) the modeled data from the background model, (e) the local auto-correlation of the observed data, and (f) the local correlation of initial data with observed data. The horizontal shows the location of the correlation functions.

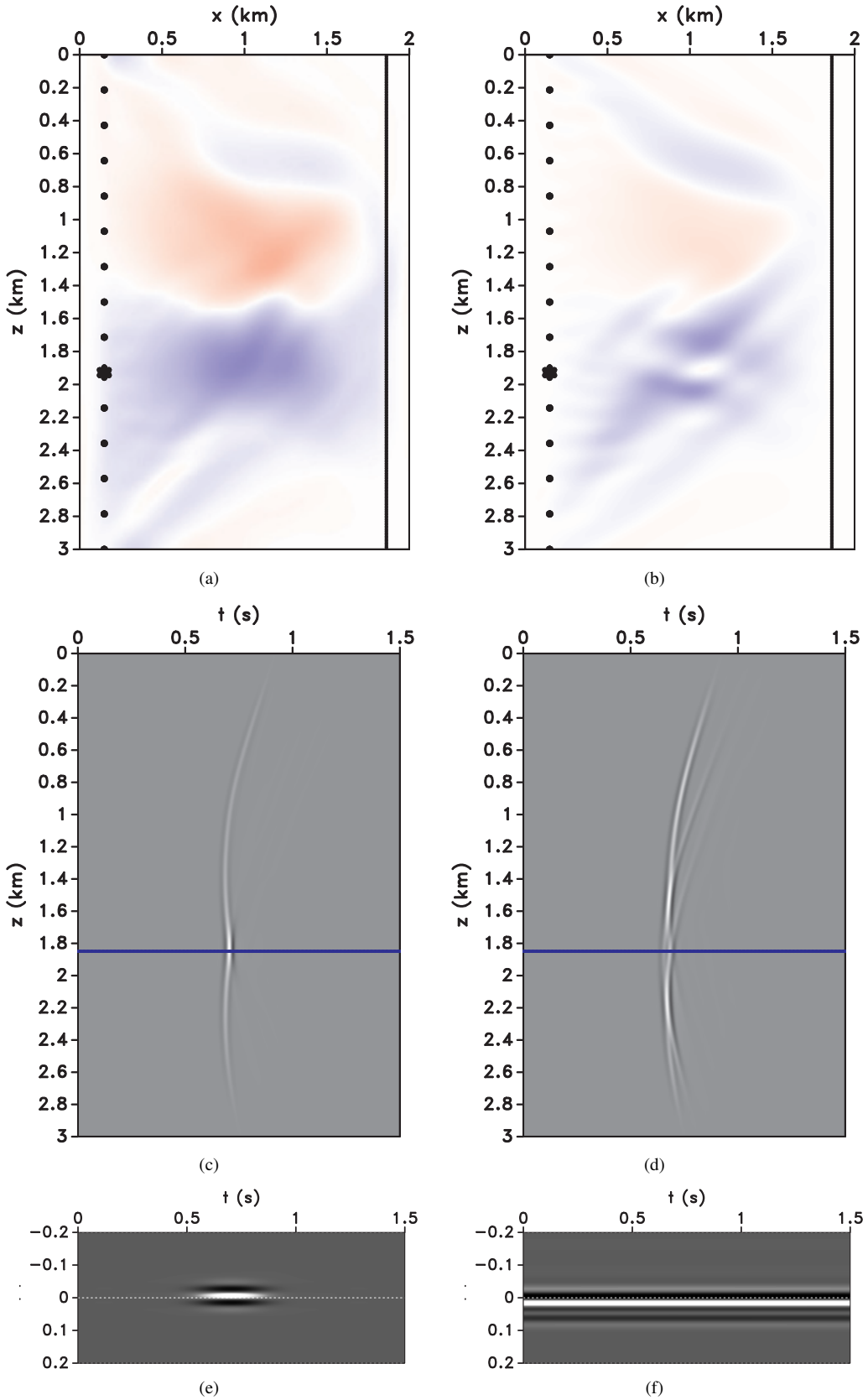


Figure 6. Inversion results: (a) shows the inverted model from local correlations, (b) the inverted model from global correlations, (c) depicts the modeled data from model in (a), (d) the modeled data from the model in (b), (e) the correlation at the final iteration for local correlation inversion, and (f) the correlation at the final iteration for the global correlation inversion. The horizontal line shows the location of the correlation functions.

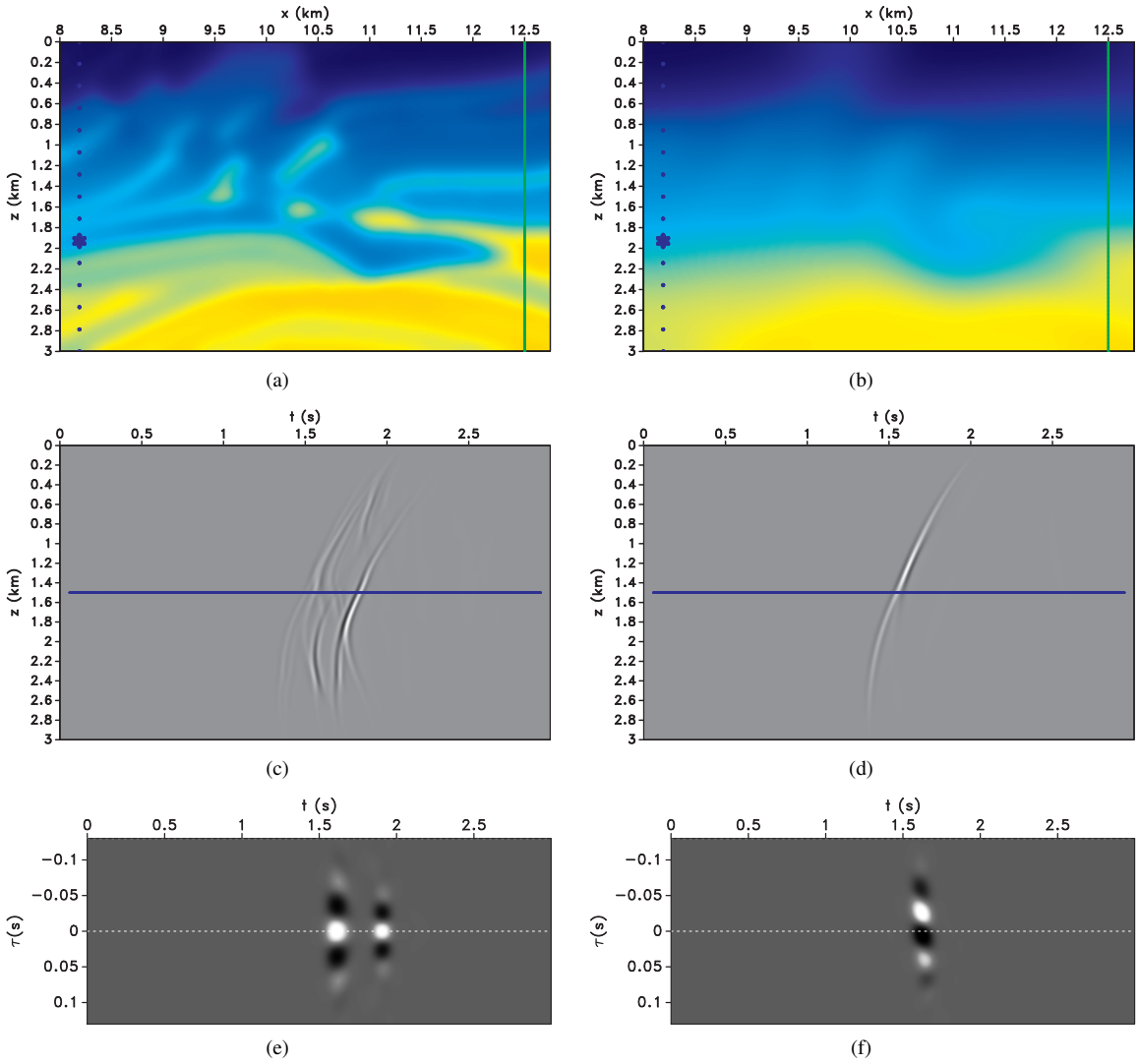


Figure 8. Marmousi cross-well experiment: (a) the true velocity model, (b) the starting velocity model, (c) the observed data, (d) the same shot gather in the initial model, (e) the local auto-correlation extracted at $z = 1.2$km, and (f) shows the local correlation from the initial data. Points in the left side of the model show the shot positions, the largest point shows the shot position for the depicted data panels. The line on the right shows the receiver array.

1982.

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.

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.

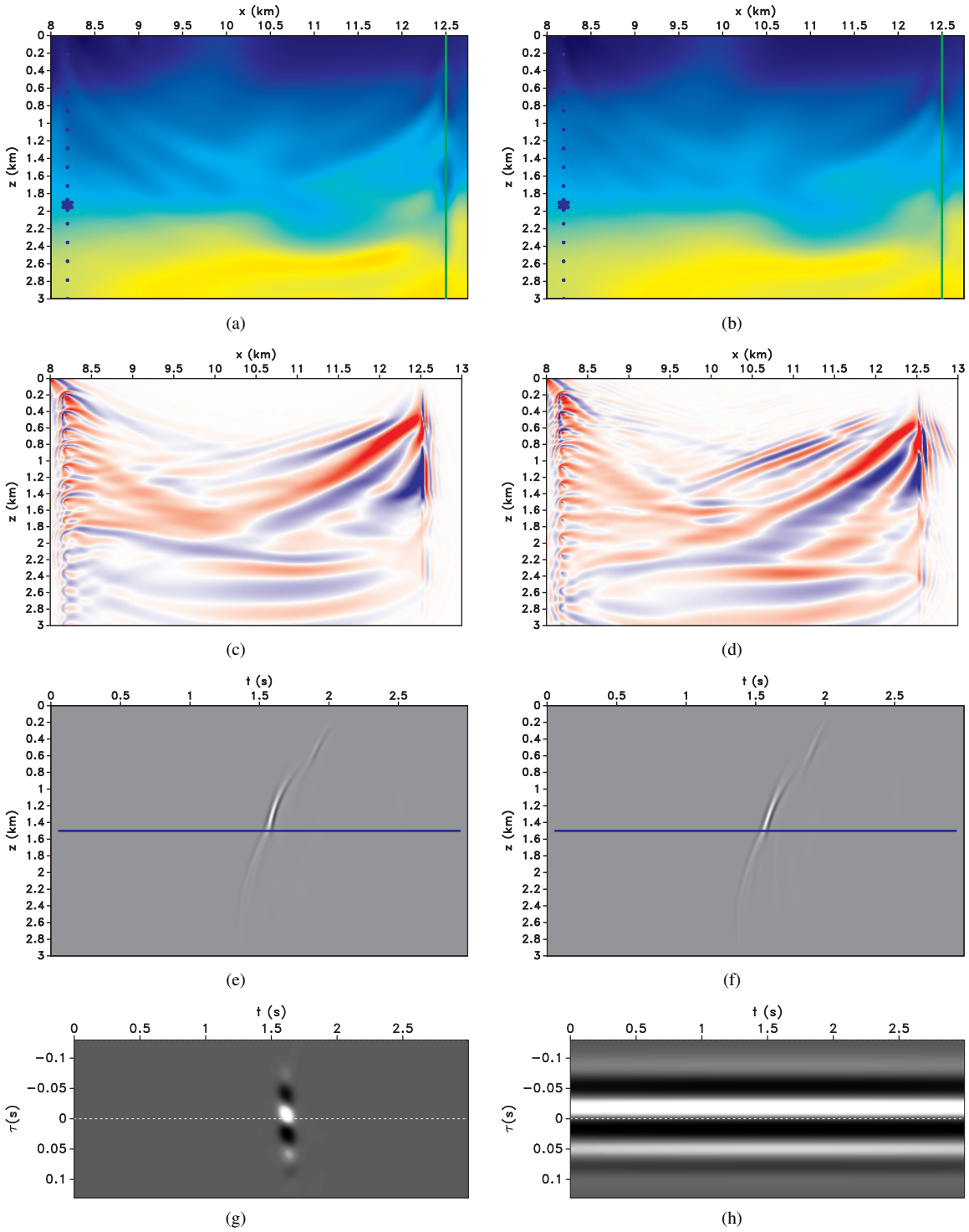


Figure 9. Inversion results from (a) local , and (b) global correlations. Panel (c) shows the initial gradient from the local correlation framework, whereas (d) depicts the initial gradient from the global correlation. The data in (e) and (f) are modeled from the local and global correlation inversion results, respectively. The bottom panels (g) and (h) show the correlation extracted at $z = 1.2$km for local and global correlations, respectively.