Full waveform inversion with an auxiliary bump functional

Pawan Bharadwaj,[1] Wim Mulder[1,2] and Guy Drijkoningen[1]

[1]*Department of Geoscience & Engineering, Delft University of Technology, Delft, The Netherlands. E-mail: bharadwaj.pawan@gmail.com*
[2]*Shell Global Solutions International B.V., Rijswijk, The Netherlands*

SUMMARY

Least-squares inversion of seismic arrivals can provide remarkably detailed models of the Earth's subsurface. However, cycle skipping associated with these oscillatory arrivals is the main cause for local minima in the least-squares objective function. Therefore, it is often difficult for descent methods to converge to the solution without an accurate initial large-scale velocity estimate. The low frequencies in the arrivals, needed to update the large-scale components in the velocity model, are usually unreliable or absent. To overcome this difficulty, we propose a multi-objective inversion scheme that uses the conventional least-squares functional along with an auxiliary data-domain objective. As the auxiliary objective effectively replaces the seismic arrivals by bumps, we call it the bump functional. The bump functional minimization can be made far less sensitive to cycle skipping and can deal with multiple arrivals in the data. However, it can only be used as an auxiliary objective since it usually does not provide a unique model after minimization even when the regularized-least-squares functional has a unique global minimum and hence a unique solution. The role of the bump functional during the multi-objective inversion is to guide the optimization towards the global minimum by pulling the trapped solution out of the local minima associated with the least-squares functional whenever necessary. The computational complexity of the bump functional is equivalent to that of the least-squares functional. In this paper, we describe various characteristics of the bump functional using simple and illustrative numerical examples. We also demonstrate the effectiveness of the proposed multi-objective inversion scheme by considering more realistic examples. These include synthetic and field data from a cross-well experiment, surface-seismic synthetic data with reflections and synthetic data with refracted arrivals at long offsets.

Key words: Numerical solutions; Inverse theory; Tomography; Non-linear differential equations; Seismic tomography.

1 INTRODUCTION

Full waveform inversion (FWI) is a nonlinear optimization procedure that estimates the Earth's model parameters by least-squares fitting of the recorded arrivals in the seismic data (Tarantola 1984; Mora 1988, 1989; Pratt 1999; Virieux & Operto 2009; Fichtner 2010). Due to the computational cost of the wave-equation modelling, the optimization is usually performed with gradient-based techniques, although several authors have tried more costly global optimization techniques (Sen & Stoffa 1991; Stoffa & Sen 1991; Sambridge & Drijkoningen 1992; Gao *et al.* 2014; Datta 2015). While fitting the observed and the modelled seismic arrivals with the conventional least-squares objective function, the gradient-based optimization will get trapped in the nearest local minimum when the error in the arrival time exceeds about half a period of the signal (Gauthier *et al.* 1986; Snieder *et al.* 1989; Mulder & Plessix 2008; Symes 2008). Here, period is related to the dominant frequency of the data. In other words, the least-squares inversion cannot recon-

struct velocity anomalies that cause shifts in the arrival times larger than half a period. Usually, the velocity anomalies accounting for the arrival times have relatively low wavenumbers. Hence, inversion of the low-frequency seismic signals with larger periods is easier and they help in the reconstruction of a kinematically correct velocity model.

1.1 Data-domain objective functions

Many authors have formulated alternative data-domain functionals to achieve global convergence in the absence of the necessary low frequencies (Luo & Schuster 1991; Shin & Min 2006; Shin & Cha 2008, 2009; Zhang & Wang 2009; van Leeuwen & Mulder 2010; van Leeuwen 2010; Bozdağ *et al.* 2011; Chauris *et al.* 2012; Donno *et al.* 2013; Warner & Guasch 2014; Engquist & Froese 2013). These can be grouped into the following two major classes.

Class 1. Functionals that give more weight to the kinematic than to the amplitude error between the seismic arrivals. They often involve the cross-correlation between the observed and modelled arrivals (Luo & Schuster 1991; van Leeuwen & Mulder 2010; van Leeuwen 2010). After cross-correlation, the arrival-time error can be picked by hand, which is tedious and to be avoided if possible. Fully automatic methods involve proper normalization of the arrivals prior to cross-correlation in order to ensure that the energy at non-zero time lags quantifies the arrival-time error. These functionals, however, suffer from cross-talk between multiple arrivals. Hence, the data are assumed to have only single arrivals or strong first arrivals. Such functionals are primarily used in cross-well data and tomographic inversion. Luo & Sava (2011) suggested replacing the cross-correlation with a deconvolution to obtain better convergence.

Class 2. Functionals that aim to fit the data after transforming them into a simpler form. By this transformation, the strong nonlinear dependence of the functional with respect to the medium parameters can be avoided (Schuster 2015). The functionals using simplified data have a basin of attraction with a larger size than that of the least-squares functional. The simplified data are easier to fit as they have *artificial* ultralow frequencies that can circumvent the cycle-skipping problem during the optimization. Creation of these artificial low frequencies by a nonlinear operation bears a similarity to other methods such as using deconvolution prior to data fitting (Fei *et al.* 2012), traveltime picking, and a differential-semblance variant that recovers missing low frequencies (Sun & Symes 2012). It should be noted that these simplifying operators act in a nonlinear way on the data, unlike the usual scale separation in the multiscale inversion approach (Bunks *et al.* 1995), which is linear. This means that they do not solely rely on the low-frequency content of the data. An example is the functional that measures the misfit between the *envelopes* of the observed and the modelled waveforms (Bozdağ *et al.* 2011; Wu *et al.* 2014; Chi *et al.* 2014; Luo & Wu 2015). The envelope operator is nonlinear and the resulting data have ultralow frequencies. Another example is the normalized integration method (Chauris *et al.* 2012; Donno *et al.* 2013) that uses the normalized time integral of the squared data for least-squares minimization. The normalized integration method, however, might suffer from noise. An advantage of using these functionals, compared to the correlation-type functionals, is that they take multiple arrivals in the data into account.

1.2 Problems using the envelope-based misfit

In this paper, we formulate another data-domain objective function that uses the data in a reduced or simplified form. We call it the *bump* functional, after the mathematical definition of a bump function. A bump function or mollifier is a smooth function in the sense of having continuous derivatives of all orders. Intuitively, given a function which is rather irregular or rough, convolution with a mollifier will *mollify* the function, that is, its sharp features are smoothed. It belongs to the second class of functionals mentioned above. The bump functional bears a strong similarity to the envelope-based misfit functional, without suffering from the following issues.

Issue 1. The envelope-based misfit may perform worse than a correlation-based functional, even in the case of a single arrival, because global convergence with a gradient-based method can only be obtained if the modelled and the observed envelopes partially overlap. In other words, this functional cannot reconstruct velocity anomalies that separate the modelled and the observed arrivals in

time by roughly more than the dominant period. This is due to the fact that the envelope-based misfit is not sensitive to arrival-time errors that exceed the overall width of the observed and modelled envelopes.

Issue 2. One of the key challenges in waveform inversion is the reconstruction of the smooth background velocity from *reflected* arrivals recorded later in time and at shorter offsets, compared to *transmitted* early arrivals. The envelope-based misfit cannot reconstruct the background velocity of the model using only reflection data. Wu *et al.* (2014) and Luo & Wu (2015) note that the envelope-based inversion results are much rougher when fitting the reflected arrivals in the data. During inversion, in order to fit only the stronger transmitted arrivals in the records, they use squared envelopes instead of just the envelopes. In this paper, we will analyse the cause of this issue in more detail.

1.3 Why the bump functional?

The bump functional can be seen as a generalized envelope-based misfit. We show that it not only can be made insensitive to cycle skipping but also has an improved global-convergence robustness compared to the envelope-based misfit. The bump functional is *sensitive* to arrival-time errors in the modelled data that are larger than a period. Furthermore, as the evaluation of the functional does not involve cross-correlation, it is also applicable to data containing multiple arrivals. The price paid is severe non-uniqueness while estimating the solution. In other words, the solution of the bump functional will depend on the initial guess and the chosen optimization method.

1.4 Importance of multi-objective strategy

In this paper, we discuss different properties of the bump functional. We note that the bump-functional inversion suffers from the same second issue as the envelope-based misfit. In order to partially overcome this difficulty, we propose a multi-objective inversion strategy using both the bump functional and the least-squares functional. We illustrate the effectiveness of this inversion strategy by a number of examples.

The rest of the paper is organized such that first we describe the conventional least-squares optimization. After that, we formulate the bump functional and demonstrate its characteristics using some numerical experiments. Finally, we describe the proposed multi-objective inversion strategy and show its effectiveness using realistic numerical examples. The last section summarizes the paper.

List of symbols

x	Horizontal coordinate
z	Vertical coordinate
x_r	Receiver horizontal coordinate
$\mathbf{x}$	Subsurface point, (x, z)
$\mathbf{x}_s$	Source position, (x_s, z_s)
$\mathbf{x}_r$	Receiver position, (x_r, z_r)
t	Time
f	Frequency
k_r	Wavenumber
f_d	Dominant frequency in observed data
τ_d	Dominant period in observed data
$\tau_{\max}$	Largest period in observed data
λ_d	Dominant wavelength at receivers
p	Modelled time-domain wavefield
q	Observed time-domain wavefield

p_b	Modelled time-domain bumpy wavefield
q_b	Observed time-domain bumpy wavefield
p_a	Modelled time-domain absolute-valued wavefield
$\tilde{q}_b$	Observed bumpy wavefield in frequency–wavenumber domain
$\tilde{p}_b$	Modelled bumpy wavefield in frequency–wavenumber domain
$\tilde{q}_a$	Observed absolute-valued wavefield in frequency–wavenumber domain
$\tilde{p}_a$	Modelled absolute-valued wavefield in frequency–wavenumber domain
ϕ	Source time function
J_{ls}	Least-squares objective function
J_b	Bump objective function
b	Blurring function
$\tilde{b}$	Blurring function in frequency–wavenumber domain
σ_t	Time-blurring parameter
σ_r	Receiver-array-blurring parameter
$\mathcal{M}$	Model space
$\mathcal{D}$	Data space
$\mathcal{D}_b$	Bumpy-data space
$\mathcal{F}$	Forward map to model wavefields
$\mathcal{F}_b$	Forward map to create bumpy data
c	Wave speed
ρ	Mass density

2 REVIEW OF THE LEAST-SQUARES INVERSION

The classic least-squares inversion aims to find a model in $\mathcal{M}$, a set of possible models of the subsurface, which minimizes the functional (Tarantola 1984),

$$J_{\mathrm{ls}} = \frac{1}{2} \sum_s \sum_r \sum_t [p(\mathbf{x}_r, \mathbf{x}_s, t) - q(\mathbf{x}_r, \mathbf{x}_s, t)]^2 . \tag{1}$$

Here, $q(\mathbf{x}_r, \mathbf{x}_s, t)$ is the observed pressure for a source at position $\mathbf{x}_s$ and receiver at $\mathbf{x}_r$ as a function of time t. The modelled data are denoted by $p(\mathbf{x}_r, \mathbf{x}_s, t)$. Both q and p belong to the data space denoted by $\mathcal{D}$. The operator $\mathcal{F} : \mathcal{M} \to \mathcal{D}$ requires the solution of a wave equation to provide data in the data space $\mathcal{D}$ for a given model in $\mathcal{M}$. For the examples in this paper, we will consider the 2-D acoustic wave equation

$$\frac{1}{\rho c^2} \frac{\partial^2 p}{\partial t} - \nabla \cdot \frac{1}{\rho} \nabla p = \phi(t) \delta(\mathbf{x} - \mathbf{x}_s), \tag{2}$$

with pressure $p(\mathbf{x}, \mathbf{x}_s, t)$, sound speed $c(\mathbf{x})$, mass density $\rho(\mathbf{x})$ and source signature $\phi(t)$. In 2-D, $\mathbf{x} = (x, z)$. We use a time-domain staggered-grid finite-difference code to model the pressure wavefield required to perform the forward and the adjoint wavefield computations (Tarantola 1984; Fichtner 2010). Absorbing boundary conditions are used on all sides of the computational domain except for a realistic example, where reflections from the boundary at $z = 0$ are included. We assume that the source signature, $\phi(t)$, is known in our examples. In all our examples, we perform full-bandwidth inversion as opposed to the multiscale inversion approach (Bunks *et al.* 1995) that relies on the low-frequency information in the data. The optimizations are performed by the conjugate-gradient method. The gradient at each iteration is preconditioned by the type 1 migration weights of Plessix & Mulder (2004).

3 THE BUMP FUNCTIONAL

3.1 Definition

To compute the bump functional, the modelled and the observed data should be mapped into the space $\mathcal{D}_b$ of *bumpy* data. The observed and modelled data in bumpy-data space are denoted by p_b

and q_b, respectively. A function, $\mathcal{F}_b : \mathcal{D} \to \mathcal{D}_b$, maps the modelled data into the bumpy-data space and is given by

$$p_b(\mathbf{x}_r, \mathbf{x}_s, t) = \mathcal{F}_b[\sigma_t, \sigma_r, p, \epsilon](\mathbf{x}_r, \mathbf{x}_s, t) = b(\mathbf{x}_r, \sigma_r) *_r$$
$$\times \left[b(t, \sigma_t) *_t \sqrt{p^2(\mathbf{x}_r, \mathbf{x}_s, t) + \epsilon^2} \right]. \tag{3}$$

The non-zero constant ϵ makes p_b differentiable at $p = 0$. We chose ϵ much smaller than p, such that $\sqrt{p^2(\mathbf{x}_r, \mathbf{x}_s, t) + \epsilon^2} \simeq |p(\mathbf{x}_r, \mathbf{x}_s, t)|$. In the above equation, b is a blurring function for which we choose a Gaussian with standard deviation equal to either σ_t when blurring in time or σ_r when blurring in one of the receiver coordinates. Convolution in time is denoted by $*_t$, and along the receiver position, $\mathbf{x}_r$, by $*_r$.

The nonlinear mapping $\mathcal{F}_b$ replaces the arrivals in the data with *bumps* and is non-injective due to the following reasons:

(i) $p_b \geq 0$, irrespective of the polarity of arrivals in p,
(ii) depending on the amount of blurring, p_b is less dependent on the source signature and
(iii) the blurring operation removes the details from p, depending on the blurring parameters, σ_t and σ_r.

The blurring parameters control the amount of blurring applied to the absolute-valued data and determine the width of the resulting bumps. For example, choosing $\sigma_t = \tau_d$ and $\sigma_r = 0$, where $\tau_d = 1/f_d$ is the period of the data corresponding to the dominant frequency f_d, would blur an impulsive arrival in the absolute-valued data such that it is spread roughly over $2\tau_d$ in time. Similarly, choosing $\sigma_t = 0$ and $\sigma_r = \lambda_d$, where $\lambda_d = c_r/f_d$ is the wavelength corresponding to the velocity, c_r, close to the receivers, would blur an impulsive arrival in the absolute-valued data such that it is spread roughly over receivers within a distance of $2\lambda_d$.

The bump functional, J_b, is the least-squares difference between the observed and the modelled bumpy data:

$$J_b = \frac{1}{2} \sum_s \sum_r \sum_t [p_b(\mathbf{x}_r, \mathbf{x}_s, t) - q_b(\mathbf{x}_r, \mathbf{x}_s, t)]^2 . \tag{4}$$

Inversion based on J_b produces a model in $\mathcal{M}$ that minimizes the difference between p_b and q_b. Details on the gradient computation of the bump functional with respect to the medium parameters are given in the Appendix. Note that minimizing their difference does not necessarily reduce the difference between p and q, since the bump functional is insensitive to the polarity of the events in the data. Also note that the bump functional with the parameters $\sigma_t/\tau_d = 0.5$ and $\sigma_r = 0$ closely resemble the envelope-based misfit (Bozdağ *et al.* 2011).

3.2 Characteristics

The main characteristics of inversion with the bump functional are listed next.

3.2.1 Insensitivity to cycle skipping

The bump-functional inversion is less sensitive to cycle skipping even in the absence of low frequencies in the data. This is due to the fact that it fits the bumpy data, as in eq. (4), which are non-oscillatory if sufficient blurring is applied.

3.2.2 Global-convergence robustness

While computing the bump functional, the absolute-valued data are blurred in time and/or along a receiver coordinate, as in eq. (3). The

role of the blurring is to increase the size of the basin of attraction of the objective function compared to that of the envelope-based misfit. The global-convergence robustness of J_b inversion, when using gradient-based optimization, increases with the amount of blurring applied to the absolute-valued data. Intuitively, after taking the absolute value, a single pulse in the modelled data will still not be able to sense a corresponding pulse in the observed data if they are separated in time. Large amounts of blurring increases the width of the observed and modelled bumps, so that they can sense each other in time during the inversion. In other words, blurring is essential to make the bump functional sensitive to large arrival-time errors. Since the bumpy data become simpler with increased blurring, it is easier to find models that fit them.

3.2.3 Non-uniqueness

The price paid for blurring is a loss of resolution and an increase of non-uniqueness. Since the mapping $\mathcal{F}_b : \mathcal{D} \to \mathcal{D}_b$ is non-injective, the inverse problem of estimating a model that minimizes the bump functional suffers from non-uniqueness, which gets worse with the amount of blurring applied in the eq. (3). Note that the least-squares inversion also suffers from non-uniqueness due to band-limitation of the data and lack of illumination, making regularization necessary to obtain an acceptable solution. The bump functional inversion suffers from non-uniqueness even when the same regularization as for the least-squares inversion is used.

Once the bump functional is minimized, the observed and modelled bumpy data match each other. However, this does not imply that the modelled data fit the observed data very well. The modelled arrivals can have the opposite polarity as well as a different wavelet, compared to the observed arrivals. Furthermore, the modelled data can contain ancillary *false* arrivals that are not present in the observed data. This prevents the bump-functional inversion from correctly updating the background velocity while fitting the

reflected arrivals. The bump functional inversion might result in a model with many high-wavenumber artefacts that generate the false arrivals. Next, we will discuss some simple examples to illustrate these characteristics.

3.3 Illustrative examples using the bump functional

3.3.1 Two arrivals

As a very simple example, we assume that both the observed and the modelled data to contain a single record with two arrivals. For the first arrival, $w_1(t)$, we use a Ricker wavelet with a peak frequency of 20 Hz and a band-limitation of 10–15–50–60 Hz. We used the same Ricker wavelet with a band-limitation of 15–17–23–25 Hz for the second arrival, $w_2(t)$. The arrivals do not contain any low-frequency information. The modelled and 'observed' data are then chosen as

$$p(t) = w_1(t + \beta - \beta_0) + 0.6\, w_2(t + \gamma - \gamma_0),$$

$$q(t) = w_1(t - \beta_0) + 0.6\, w_2(t - \gamma_0).$$

To the observed data, we added 20 per cent of white random noise. The model parameters were chosen as $\beta_0 = 2.16$ s and $\gamma_0 = 2.55$ s. Fig. 1 plots the observed and modelled data in blue and red, respectively. The inversion parameters are the arrival-time errors, β and γ. Both should be zero at the global minimum. We now consider the inverse problem of estimating the arrival times using different objective functionals.

3.3.2 Least-squares functional

Fig. 2(a) shows the classic least-squares functional, $J_{ls} = \int [p(t) - q(t)]^2 \mathrm{d}t$, as a function of the arrival-time errors. The oscillatory nature of the arrivals causes various local minima to occur around the global minimum, shrinking the radius of the basin of attraction.
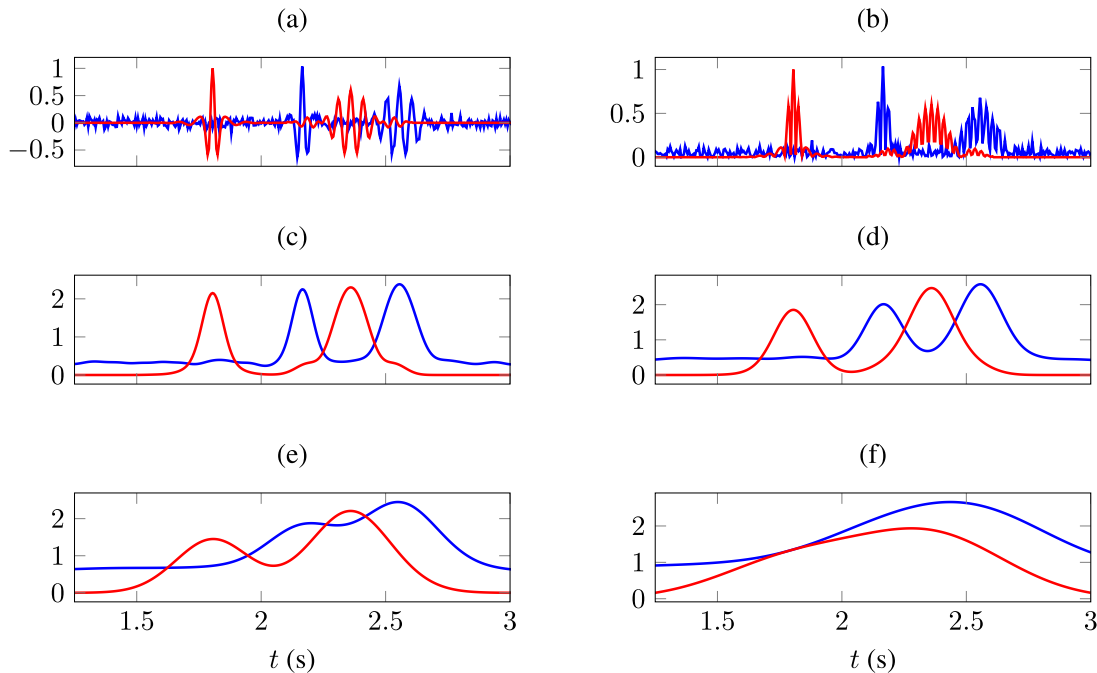
Figure 1. The observed and modelled data with two arrivals are plotted in blue and red, respectively. (a) The arrival-time error in the modelled arrivals is more than the dominant period, τ_d. (b) Absolute-valued data. (c) Bumpy data obtained using $\sigma_t/\tau_d = 0.5$, (d) $\sigma_t/\tau_d = 1$, (e) $\sigma_t/\tau_d = 2$ and (f) $\sigma_t/\tau_d = 4$.
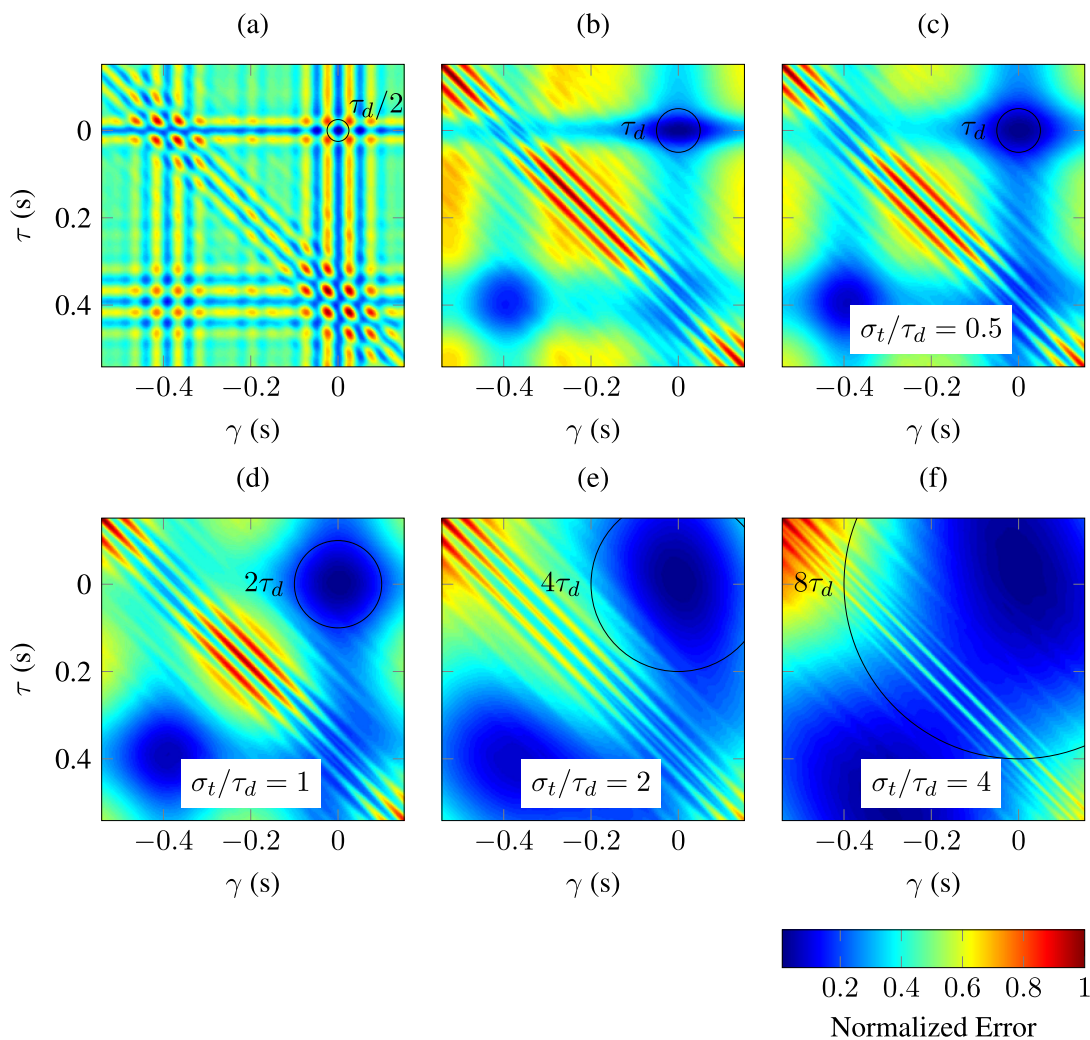
Figure 2. Various objective functions are plotted as a function of the arrival-time errors, γ and β, for the observed and modelled data shown in Fig. 1. The basin of attraction is outlined by a black circle. (a) The least-squares functional, J_{ls}. (b) The envelope-based misfit. (c) The bump functional, J_b, with $\sigma_t/\tau_d = 0.5$, (d) $\sigma_t/\tau_d = 1$, (e) $\sigma_t/\tau_d = 2$, and (f) with $\sigma_t/\tau_d = 4$.

In this case, the radius of the basin is approximately $\tau_d/2 = 0.025$ s, where, again, τ_d denotes the period of the dominant frequency. In the multiscale inversion approach (Bunks *et al.* 1995), a low-pass filter is applied to the data to first fit the lowest reliable frequencies. The radius of the basin is then half the period corresponding to the lowest frequency, $\tau_{max} = 0.1$ s. This implies that, if the arrival-time error in the initially modelled data exceeds 0.1 s, the least-squares inversion cannot accurately estimate the arrival times.

3.3.3 Envelope-based misfit

The envelope-based misfit (Bozdağ *et al.* 2011) computes the least-squares error using the observed and the modelled envelopes. In this example, Fig. 2(b) shows that the misfit suffers less from cycle skipping and local minima do not occur close to the global minimum because the envelopes are non-oscillatory. In this case, the radius of attraction roughly equals the dominant period, $\tau_d = 0.05$ s. It can be seen that the functional is insensitive to arrival-time shifts larger than τ_d, when there is no overlap between the observed and modelled data envelopes.

3.3.4 Bump functional

To compute the bump functional, the absolute-valued data are considered, as plotted in Fig. 1(b). The choice of σ_t determines the amount of blurring applied to the absolute-valued data. For $\sigma_t/\tau_d = 0.5$, the observed and modelled bumps, as plotted in Fig. 1(c), are similar to envelopes. In this case, the bump functional has characteristics similar to the envelope-based misfit, as plotted in Fig. 2(c). In general, for any given σ_t, since the observed and the modelled bumps are non-oscillatory, the bump functional suffers less from cycle skipping.

In order to corroborate our claim that the global-convergence robustness of J_b depends on the amount of blurring, we now examine the bump functional for different values of σ_t. The bumpy data for different values of σ_t are plotted in Figs 1(c)–(f). As can be seen in Figs 2(d)–(f), the radius of the attraction circle of the bump functional increases with σ_t. With increased blurring, the separated bumps of the observed and modelled data can sense each other in time, resulting in an improved global-convergence robustness. Note that the radius of the basin of attraction for J_b, unlike the least-squares functional, is independent of the presence of low frequencies in the data. The radius of basin of attraction in the case of $\sigma_t/\tau_d = 4$

(a)

(b)

(c)

(d)

t (s)
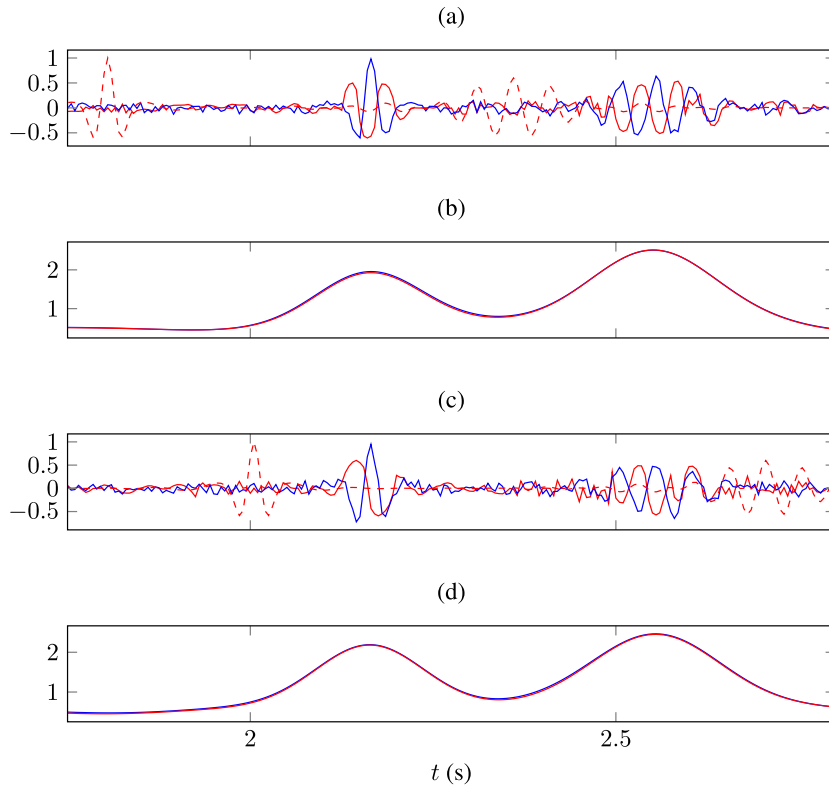
Figure 3. Plots corresponding to the two-arrivals example as in Fig. 1. Two inversions (I and II) are performed, starting from different initial values of γ and β using the bump functional with $\sigma_t/\tau_d = 1$. The observed and estimated data are plotted in blue and red, respectively. The initially estimated data are drawn with a dashed red line. (a) Data corresponding to inversion I. (b) Bumpy data corresponding to inversion I. (c) Same as (a), but for inversion II. (d) Same as (b), but for inversion II.

and 8 is larger than period corresponding to the minimum frequency in the data, $\tau_{\max}$, which shows that J_b can also be effective in the absence of low frequencies.

Finally, we show that the data minimizing the bump functional cannot be determined uniquely. We perform two inversions (I & II) using the conjugate-gradient method starting from different initial values of β and γ. During each inversion, we estimate the data that minimize the bump functional for $\sigma_t/\tau_d = 1$. We expect that the estimated data output from both the inversions to be different while fitting each other in their bumpy form. This is possible because the nonlinear mapping $\mathcal{F}_b$ is non-injective. The observed and initially estimated data for both the inversions are plotted in Figs 3(a) and (c), respectively. The estimated data after the inversions, plotted in Figs 3(a) and (c), are different and they contain false arrivals and have polarity mismatches. For both the inversions, the estimated bumpy data, plotted in Figs 3(b) and (d), at the last iteration do fit the observed bumpy data.

3.3.5 Cross-well transmission example

We now consider a cross-well experiment with a Gaussian anomaly. This is a transmission problem for which the background velocity needs to be estimated. We used evenly spaced vertical arrays of sources and receivers at $x = 1$ m and $x = 96$ m, respectively. The source wavelet had a peak frequency of 140 Hz. The assumed Earth model and the corresponding 'observed' data are plotted in Figs 4(a) and 5(a), respectively. The initial velocity model for inversion is homogeneous with $c = 1800\,\mathrm{m\,s^{-1}}$. The corresponding modelled data are plotted in Fig. 5(b) and have arrival-time errors of more

than three times the dominant period, $T = 0.0071$ s. Therefore, larger amounts of blurring are necessary during the J_b inversion to ensure that the bumps in the initially modelled data can sense the bumps in the observed data. We used the J_b functional with $\sigma_t/\tau_d = 0.5, 1, 2$ and 4. Gaussian smoothing is applied to the gradient at each iteration with a standard deviation of 1.5 m in order to avoid high-wavenumber artefacts in the solution. In all the cases, the iterations were stopped when the convergence slowed down too much.

Figs 4(b)–(e) show the velocity models and Figs 5(c)–(f) show the modelled arrivals at the last iteration. For $\sigma_t/\tau_d = 0.5$ and 1, the inversion was not able to reconstruct the velocity anomaly, as in Figs 4(b)–(c), since the J_b functional has a small basin of attraction. At the last iteration, the observed and the modelled data in bumpy form do not match each other, as mentioned in Table 1. Only the error in records with initial arrival-time error less than $2\tau_d$ is minimized (see Figs 5c–d). The J_b inversion with $\sigma_t/\tau_d = 0.5$ is similar to the envelope-based inversion, as shown in the Figs 1 and 2. In the case of $\sigma_t/\tau_d = 2$ and 4, the bump functional inversion has successfully reconstructed the velocity anomaly, as in Figs 4(d) and (e), proving that the size of the basin of attraction for J_b increases with the amount of blurring. As shown in the Table 1, the inversion is able to fit the observed bumpy data. Also, the modelled data after inversion, shown in Figs 5(e) and (f), reasonably fit the observed data in Fig. 5(a). Hence these solutions are acceptable as summarized in the Table 1.

To illustrate the non-uniqueness problem, we choose $\sigma_t/\tau_d = 8$ for J_b inversion. At the last iteration, the inversion is able to fit the observed and the modelled data in bumpy form as mentioned in Table 1. However, Fig. 5(g) shows that the modelled data do not
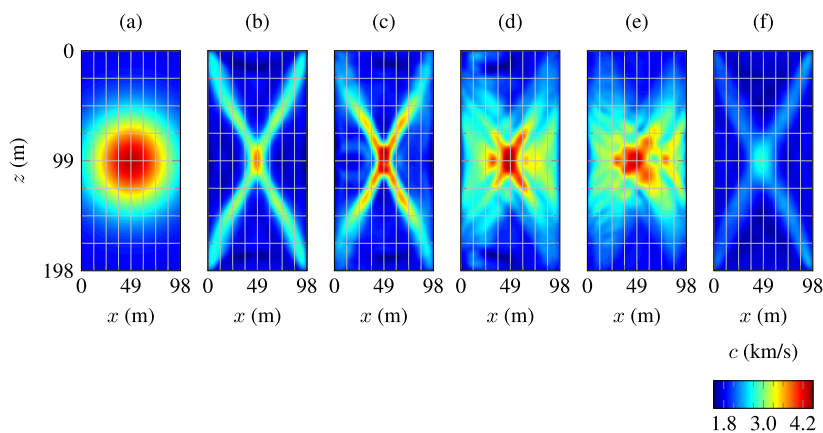
Figure 4. For the cross-well experiment with a Gaussian anomaly, J_b inversion is performed with different amounts of time blurring. The initial model before inversion is a homogeneous model with $c = 1800 \, \mathrm{m \, s^{-1}}$. (a) Earth model with a Gaussian anomaly at the centre. (b) Result using $\sigma_t / \tau_d = 0.5$, (c) $\sigma_t / \tau_d = 1$, (d) $\sigma_t / \tau_d = 2$, (e) $\sigma_t / \tau_d = 4$ and (f) $\sigma_t / \tau_d = 8$.
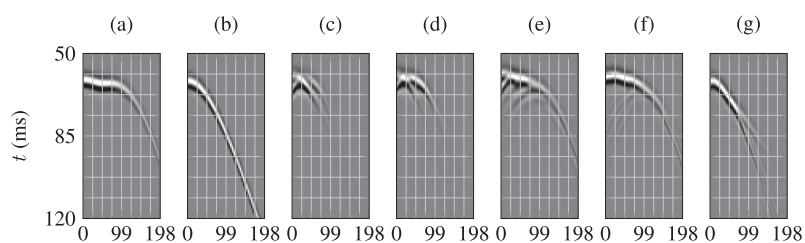
Figure 5. Data panels corresponding to the cross-well experiment in Fig. 4 for a source at (1, 0). J_b inversion is performed using different amounts of time blurring and $\sigma_r = 0$. The shot gathers at the last iteration are plotted with the same grey scale. (a) Observed shot gather. (b) Initially modelled shot gather. (c) Using J_b with $\sigma_t / \tau_d = 0.5$, (d) with $\sigma_t / \tau_d = 1$, (e) with $\sigma_t / \tau_d = 2$, (f) with $\sigma_t / \tau_d = 4$ and (g) with $\sigma_t / \tau_d = 8$.

Table 1. Convergence of the bump functional for different amounts of blurring in the case of a cross-well experiment on a Gaussian anomaly.

σ_t / τ_d	p_b fits q_b	p fits q	Comment
0.5	no	no	Lack of sensitivity
1	no	no	Lack of sensitivity
2	yes	yes	Acceptable solution
4	yes	yes	Acceptable solution
8	yes	no	Non-uniqueness

necessarily fit the observed data. There are many models providing data that can fit the observed data in their bumpy form. Since the non-uniqueness becomes more severe with larger amounts of blurring, the inversion fails to reach an acceptable solution, as shown in Fig. 4(f). Also, false arrivals appear in the modelled data and they assist in the minimization of the bump functional.
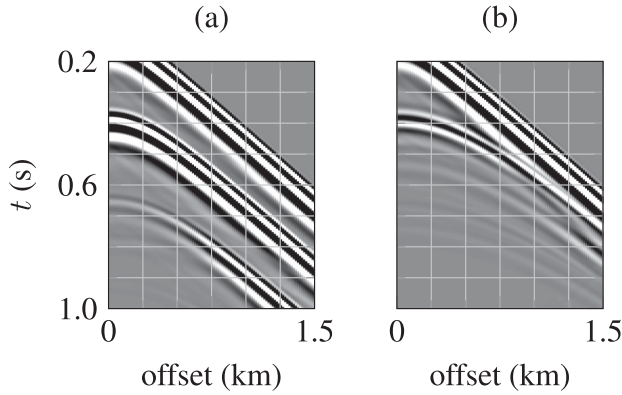
3.3.6 Three-layer reflection example

We consider another example in which the 'observed' data contain primarily reflected arrivals. In the case of reflection problems, both the background velocity and the position of the reflectors have to be estimated. The background velocity is determined by the offset-based moveout information in the data. This example illustrates the inability of the bump functional to estimate the background velocity using reflected arrivals. We also show that the velocity model that minimizes the bump functional cannot be uniquely determined.

We consider an assumed Earth model of 1500 m width and 600 m depth with a negative velocity anomaly of approximately 40 per cent. The velocity model, plotted in Fig. 6, is assumed to only vary with depth and contains two reflectors at $z = 120$ m and $z = 320$ m. We used a horizontal array of 100 evenly spaced receivers at a depth of 30 m and a source at (0, 20) m. We generated the data using a fourth-order minimum-phase Butterworth wavelet of
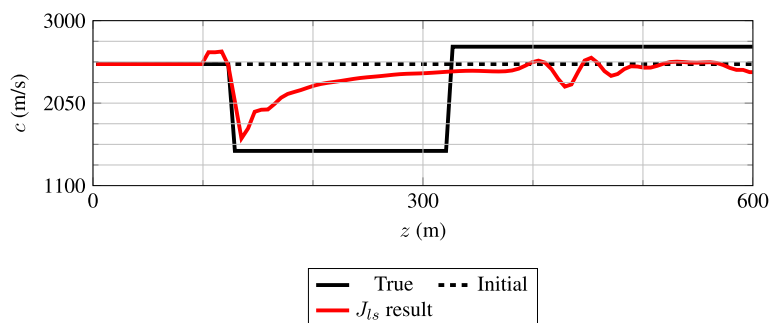
Figure 6. Results of single-objective J_{ls} inversion of the three-layer reflection example. The initial model (dashed black) before inversion is also shown.

Figure 7. The shot gathers corresponding to the three-layer reflection experiment in Fig. 6. (a) Observed gather. (b) Cycle-skipped modelled gather at the last iteration of the J_{ls} inversion.

bandwidth 10–30 Hz. Due to the negative velocity anomaly, head waves do not arrive at the receivers before 1.2 s recording time. Fig. 7(a) depicts the 'observed' shot gather, where an internal multiple can also be seen. The initial model for inversion is homogeneous with $c = 2500\,\mathrm{m\,s}^{-1}$. The spatial sampling for the inversion mesh is 6 m.

Fig. 6 shows the result of the J_{ls} inversion. It fails to reconstruct the background velocity of the model and convergence starts to become very slow after 20 iterations. After updating the reflectivity of the first layer, the J_{ls} inversion updates the reflectivity of the model at depths between 400 and 450 m. This causes the modelled data, plotted in Fig. 7(b), to partially match the 'observed' data, primarily at the short offsets. We also see that these reflectors, positioned at the wrong depths, generate arrivals in the data that are cycle-skipped at larger offsets. This behaviour of least-squares inversion is well known. The cycle skipping indicates that the solution is caught in a local minimum. It has to be noted that the results might differ depending on the coarseness of the inversion mesh (Ma *et al.* 2012). We did neither apply smoothing to the gradient at each iteration nor use any additional smoothness constraints on the model. In the case of this simple example, we noted that the least-squares inversion was able to reach the global minimum when the spatial sampling for the inversion mesh was 10 m.

We have seen that the bump functional does not suffer from cycle skipping if enough blurring is applied. In order to see if it can update the background velocity, we performed the bump-functional inversion with $\sigma_t/\tau_d = 0.5$, 1, 2 and 4. No blurring along the receiver coordinate is applied since this is an illustrative
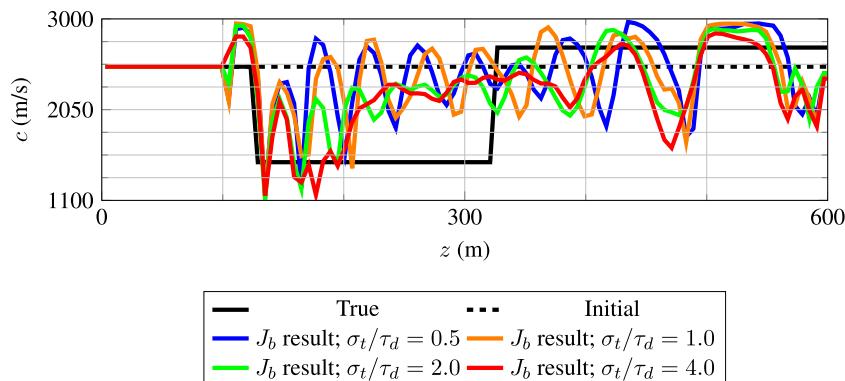
example. The observed bumpy data for these cases are plotted in Figs 9(e)–(h). The modelled bumpy data at the final iteration are displayed in Figs 9(i)–(l). We note that the inversion has matched the observed and the modelled data in their bumpy form at the last iteration. However, the resulting velocity models have an incorrect background velocity, as Fig. 8 shows. The J_b inversion has not updated the background velocity but only boosted the reflectivity of the model at depths between 400 and 550 m. In all cases, the modelled data contain a lot of *false* arrivals and do not fit the observed data, as can be seen in Figs 9(a)–(d). These false arrivals help the J_b inversion in matching the bumpy form of the observed and modelled data. The false arrivals are caused by several reflectors at incorrect depths. Even in the presence of moveout information in the observed data, we see that the bump functional cannot update the background velocity of the model.
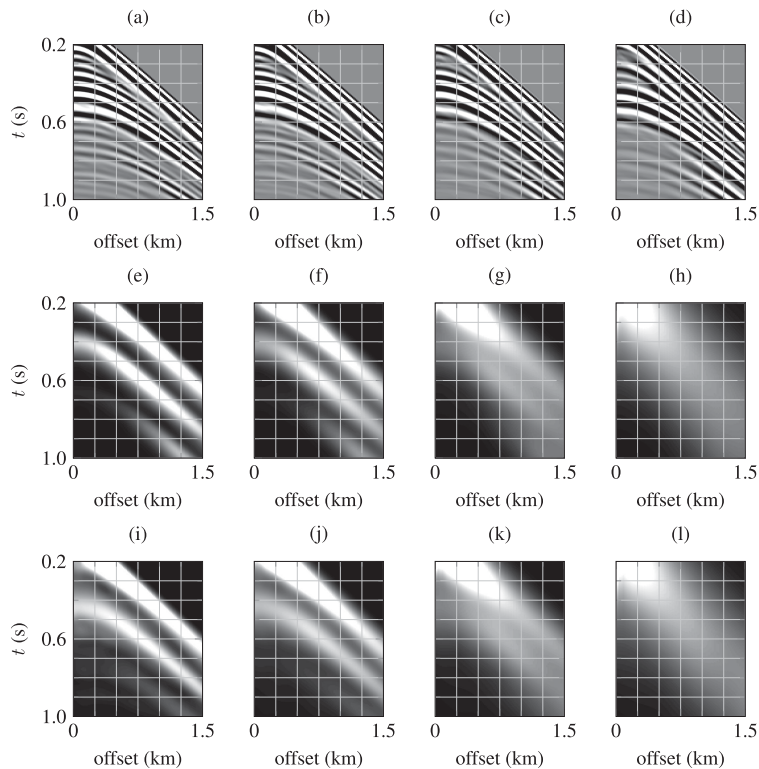
We now perform the bump functional inversion with $\sigma_t/\tau_d = 1$ and $\sigma_r = 0$ starting from two different initial models (I and II), plotted in Fig. 10. The output models in both the cases are different in a non-trivial way since the data corresponding to them, plotted in Figs 11(a) and (b), do not match each other. However, the bumpy data corresponding to the output models, plotted in Figs 11(c) and (d), are similar. This shows that the velocity model that explains the observed-bumpy reflections is not unique. Due to this, the bump functional inversion fails to output a model with the correct background velocity, even though it suffers less from cycle-skipping.

4 MULTI-OBJECTIVE INVERSION

As mentioned before, least-squares inversion suffers from cycle skipping and cannot recover velocity errors that cause arrival-time errors larger than $\tau_d/2$. Let $\mathcal{M}_{ls}$ denote a set that includes all the models corresponding to the local minima of the least-squares objective function, excluding the global minimum. We assume that the bump functional does not suffer from cycle skipping, for particular values of σ_t and σ_r, and let $\mathcal{M}_b$ denote a set that contains all the non-unique solutions that minimize it. We state without a proof that the sets $\mathcal{M}_{ls}$ and $\mathcal{M}_b$ are disjoint due to the following intuitive reasons:

(i) A model that belongs to $\mathcal{M}_{ls}$ generates data that are cycle-skipped compared to the observed data, but this model does not belong to $\mathcal{M}_b$ since the bump functional does not suffer from cycle skipping for sufficiently large blurring.

(ii) A model that belongs to $\mathcal{M}_b$ generates data that contain false arrivals, but this model does not belong to $\mathcal{M}_{ls}$ because false arrivals increase the least-squares functional.

Figure 8. Same as Fig. 6, except that the bump functional is used with different amounts of time blurring and $\sigma_r = 0$.

Figure 9. Modelled shot gathers after J_b inversion corresponding to the reflection experiment in Fig. 8. No blurring is applied along the receiver array ($\sigma_r = 0$) in all cases. (a) Data using $\sigma_t/\tau_d = 0.5$, (b) $\sigma_t/\tau_d = 1$, (c) $\sigma_t/\tau_d = 2$ and (d) $\sigma_t/\tau_d = 4$. (e) Observed bumpy data using $\sigma_t/\tau_d = 0.5$, (f) $\sigma_t/\tau_d = 1$, (g) $\sigma_t/\tau_d = 2$ and (h) $\sigma_t/\tau_d = 4$. (i) Modelled bumpy data after J_b inversion using $\sigma_t/\tau_d = 0.5$, (j) $\sigma_t/\tau_d = 1$, (k) $\sigma_t/\tau_d = 2$ and (l) $\sigma_t/\tau_d = 4$.
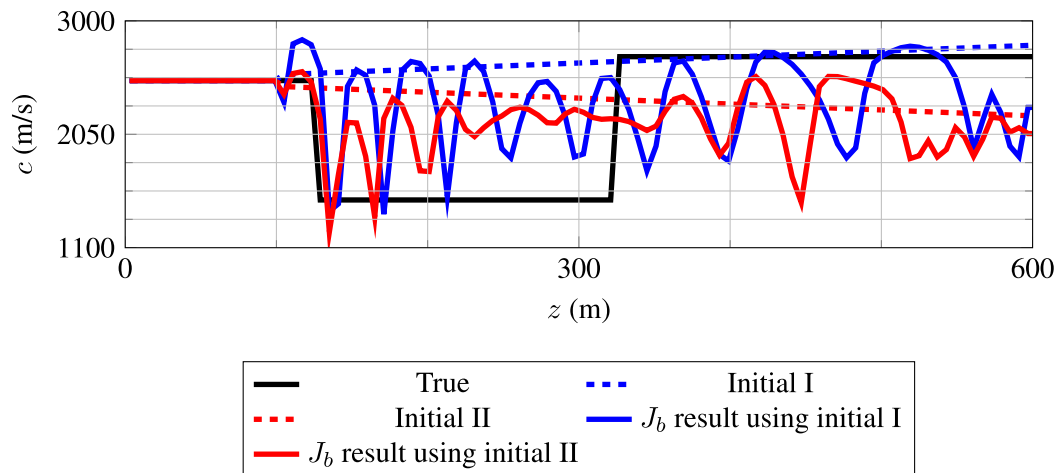
Figure 10. Same as Fig. 6, except that the bump functional is used with $\sigma_t/\tau_d = 1$ and $\sigma_r = 0$. Two different initial models were used for inversion.

This motivates the use of both the least-squares and the bump functional in the inversion. In the inversion strategy, the primary objective is to minimize the least-squares difference. It is also used to constrain the model space $\mathcal{M}_b$. Minimization of the bump functional is an auxiliary objective needed to move away from models that belong to $\mathcal{M}_{ls}$.
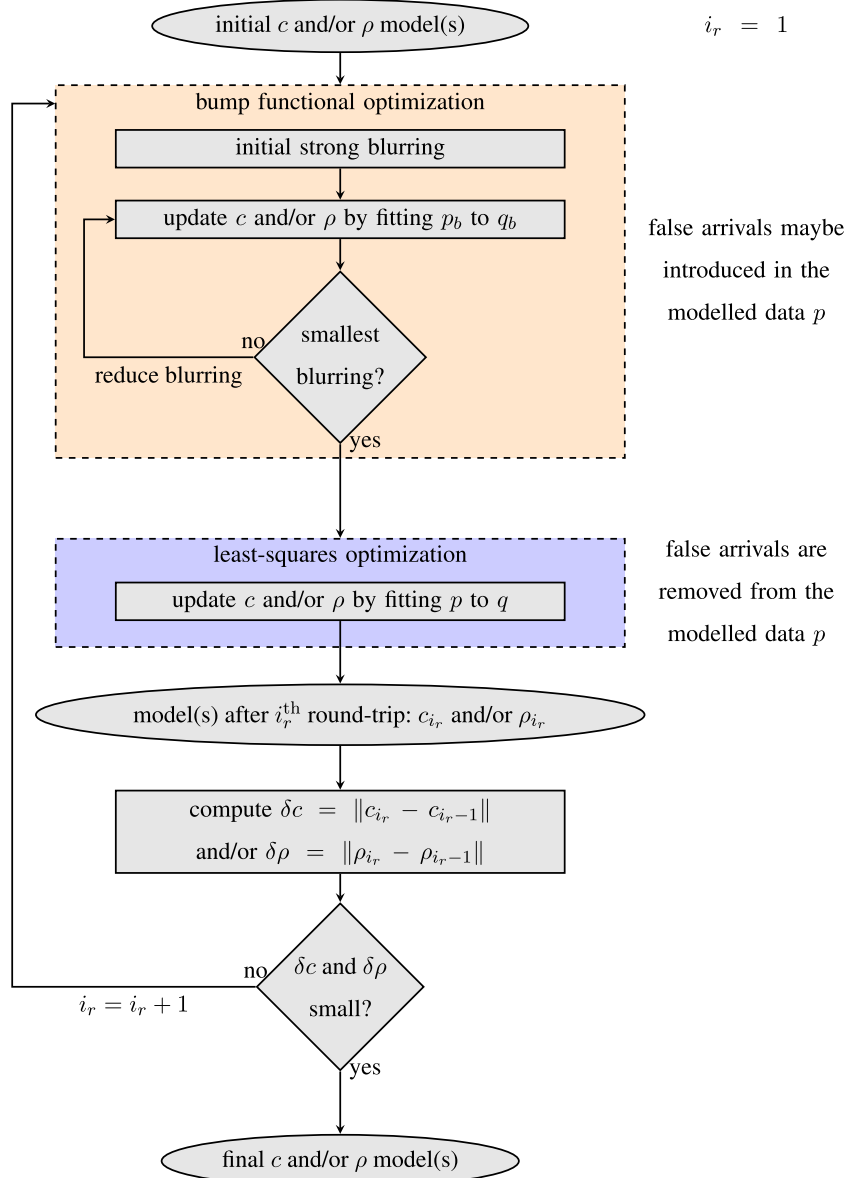
4.1 Strategy

The multi-objective inversion scheme we use throughout this paper is given in Fig. 12. The inversion consists of several round-trips depending on how far the starting model is from the global minimum.

Within each round-trip, we optimize both the J_b and J_{ls} objectives separately. During each individual optimization, we update the subsurface models and the iterations are stopped whenever convergence becomes too slow. The multi-objective inversion stops when the change in the output between consecutive round-trips is negligible. This implies that the inversion has converged to a model that simultaneously minimizes the least-squares and the bump functional.

We start with the bump-functional inversion, where strong blurring is chosen both in time and along the receiver coordinate. Then, we gradually reduce the blurring and, at the same time, perform more iterations. The motivation for starting with strong blurring is
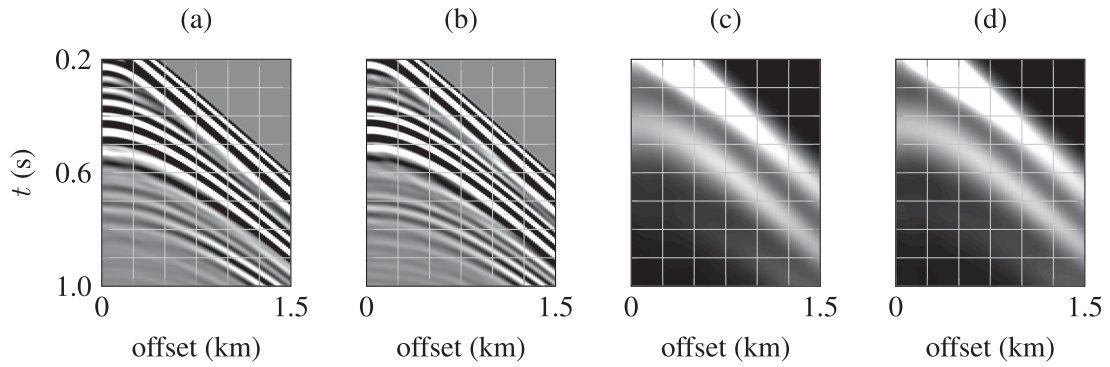
Figure 11. Single-objective bump functional inversion using $\sigma_t/\tau_d = 1$ and $\sigma_r = 0$ starting from two different initial models (I and II), as plotted in Fig. 10. (a) Modelled data at the last iteration using initial model I and (b) initial model II. (c) Modelled bumpy data at last iteration using initial model I and (d) initial model II.

Figure 12. Flowchart of multi-objective inversion strategy.

Table 2. Blurring parameters used in different examples during the multi-objective inversion. Strong blurring is chosen initially depending on the complexity of the example. The blurring is then reduced to the lowest as shown in Fig. 12. The multi-objective inversion is not sensitive to the exact blurring parameters chosen.

Example	Initial strong blurring $(\sigma_t/\tau_d, \sigma_r/\lambda_d)$	Reduced lowest blurring $(\sigma_t/\tau_d, \sigma_r/\lambda_d)$
Three-layer reflection	skipped	(1, 0)
Cross-well with field data	(1.5, 1.0)	(0.5, 0)
Five layer: early arrivals	(1.0, 0.5)	(0.5, 0)
Five layer: all arrivals	(1.5, 1.0)	(0.5, 0)
Complex 2-D model	(1.0, 0.5)	(0.5, 0)

that it allows the observed and modelled bumps, if they are far apart, to interact with each other so that the time gap between them can be closed. Once the bumps are close enough, inversion with a smaller amount of blurring will provide a fit with better resolution. Instead of choosing strong initial blurring, the amount of blurring can also be chosen according to the complexity of the problem and hence by considering the initially modelled data. A strong initial blurring, in time and along the receiver coordinate, can be used only if the arrival-time errors are large in the initial comparison between the modelled and observed data. Otherwise, if the arrival-time errors are low, the data are blurred either in time or along the receiver coordinate using a small amount of initial blurring. This will reduce the computational burden of the multi-objective inversion. Note that the purpose of blurring is to just create an overlap between the modelled and the corresponding observed bumpy arrivals. As long as this overlap is created, the exact choice of blurring parameters will not affect the multi-objective inversion. The resulting model that minimizes the bump functional is non-unique and its corresponding data will contain false arrivals. We therefore use the output of the bump-functional optimization as input to the least-squares optimization to complete a round-trip. The least-squares objective removes the false arrivals that are not present in the observed data. However, it could happen that the least-squares optimization now converges to a local minimum. Then we use the output corresponding to the, possibly local, minimum of the least-squares inversion from the first round-trip as input to the bump-functional optimization. The auxiliary bump objective now pulls the solution out of the local minimum, since it does not suffer from cycle skipping. Subsequently, we carry out more round-trips to converge to the global minimum of the least-squares objective.

4.2 Three-layer reflection example, again

To demonstrate the effectiveness of our multi-objective inversion scheme, we consider the same three-layer reflection example with a negative velocity anomaly as used before for single-objective inversion with either J_{ls} or J_b. In each round-trip, we first use the bump functional, followed by minimization of the least-squares functional, as shown in the Fig. 12. The blurring parameters used in this case are given in the Table 2. For this simple example, we did not use the bump functional with strong blurring. The final model after 9 round-trips is plotted in green in Fig. 13. The corresponding modelled shot gather, plotted in Fig. 14(f), is not cycle-skipped when compared to the observed shot gather in the Fig. 7(a).

During the first round-trip, the bump functional inversion outputs a model that belongs to $\mathcal{M}_b$ as shown in Fig. 13. The modelled data corresponding to this model, in Fig. 14(a), has false arrivals due to strong reflectivity around $z = 450$ m. We note that the J_{ls} optimization removes those false arrivals that are not present in the observed data, as visible in Fig. 14(b). The J_{ls} inversion converges to a local minimum, where the modelled data are cycle-skipped compared to the observed data. We now use the bump functional to pull the trapped solution out of the local minimum during the second round-trip. The output model and the corresponding modelled data are plotted in Figs 13 and 14(c), respectively. We note that whenever the least-squares objective removes the false arrivals, it updates the background velocity of the model. J_{ls} inversion again converges to another local minimum that is closer to the global minimum than after the first round-trip. This behaviour is reminiscent of Mora's (1989) observation that the J_{ls} functional can provide low-wavenumber updates to the velocity model, provided the reflectivity is strong enough to allow for an interaction between the scattered and the direct wavefield. In our case, after each round-trip, the output of the least-squares inversion suffers less and less from cycle skipping and provides convergence to the global minimum.

5 REALISTIC EXAMPLES

### 5.1 Cross-well example with field data

We consider a cross-well experiment to demonstrate the applicability of the multi-objective inversion strategy to field data. The field data were also used by van Leeuwen (2010) and van Leeuwen & Mulder (2010) to perform wave-equation based traveltime
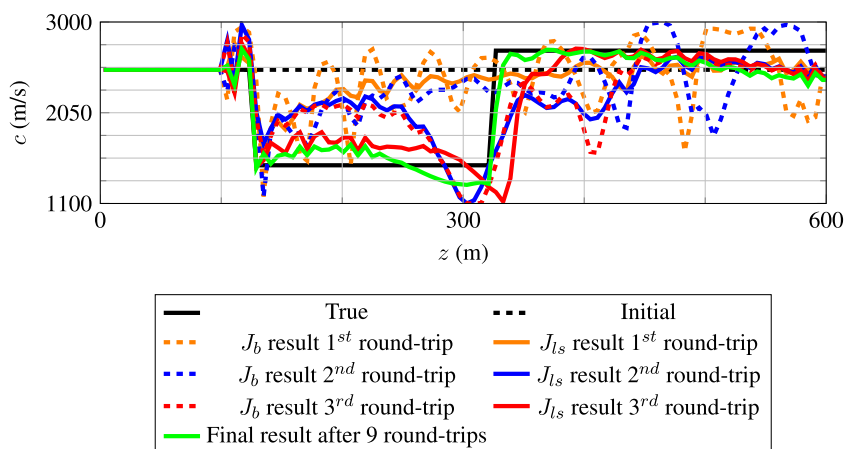
Figure 13. Same as Fig. 6, but now with the proposed multi-objective inversion scheme. The bump functional with $\sigma_t/\tau_d = 1.0$ and no receiver blurring ($\sigma_r = 0$) is used along with the least-squares objective.
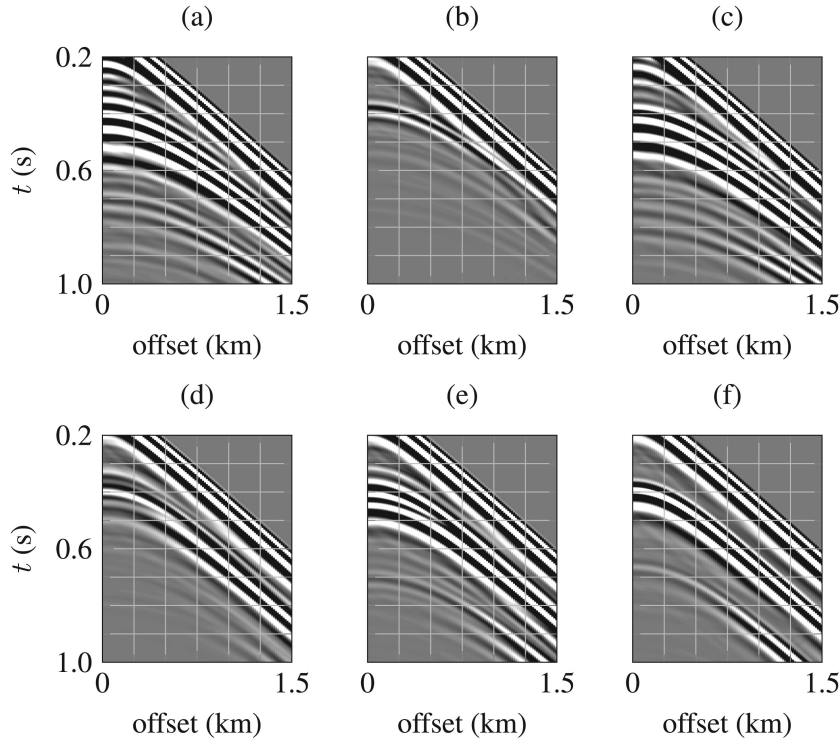
Figure 14. Modelled shot gathers at the last iteration corresponding to the multi-objective inversion in Figs 13. (a) After the first J_b inversion. (b) After the first J_{ls} inversion. (c) After the second J_b inversion. (d) After the second J_{ls} inversion. (e) After the third J_b inversion. (f) After the third J_{ls} inversion.
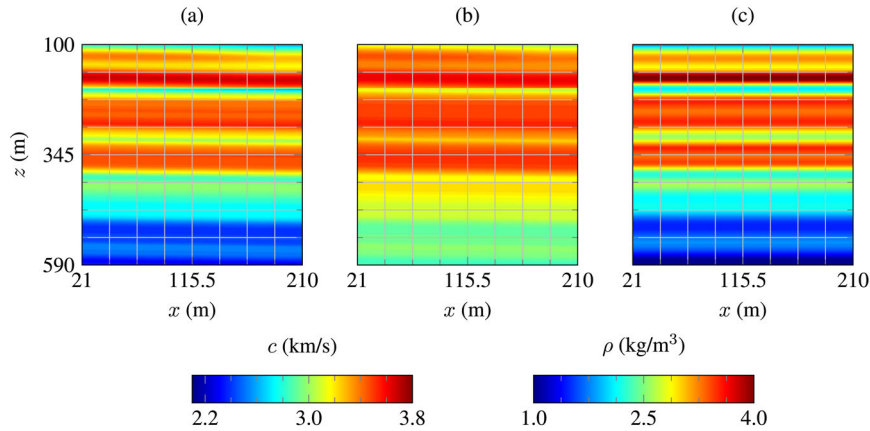

Figure 15. A synthetic cross-well example. (a) c and (b) ρ models after interpolating the well-logs used for forward modelling. (c) Multi-objective inversion output c model using synthetic data from the well-log models. The output c and ρ models are only allowed to vary with depth. The ρ model is not shown.

tomography with a cross-correlation type functional. Two wells are located at $x = 205$ m and $x = 27$ m with 122 sources and 125 receivers, respectively. The source and receiver spacing interval is approximately 3.84 m. The velocity and density Earth models after interpolating the well-log data are plotted in Figs 15(a) and (b). During the inversion, both velocity and density are estimated and a 110-Hz Ricker source wavelet is used. We choose a homogeneous initial velocity model with $c = 2500$ m s^{-1} for inversion such that the arrival-time error in the initially modelled data, plotted in Fig. 16(c), exceeds the dominant period by a factor 2 to 3. Due to this reason, J_b inversion with high blurring is used initially in the multi-objective inversion strategy, which is outlined in the Fig. 12. The blurring parameters used during the multi-objective inversion are given in

Table 2. It is expected that the single-objective J_{ls} inversion suffers from cycle-skipping.

5.1.1 Synthetic data inversion

We first use the interpolated well-log models to generate synthetic pressure data, as plotted in Fig. 16(a). While inverting these data, the output models are only allowed to vary with depth. The bumpy 'observed' data used during the inversion are plotted in Figs 17(a) and (c). The multi-objective inversion results in the final velocity model plotted in Fig. 15(c). The modelled data after the final iteration, plotted in Fig. 16(d), match the synthetic observed data. And
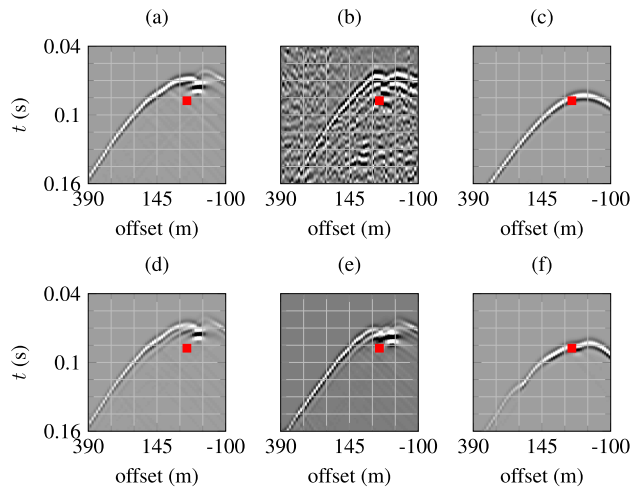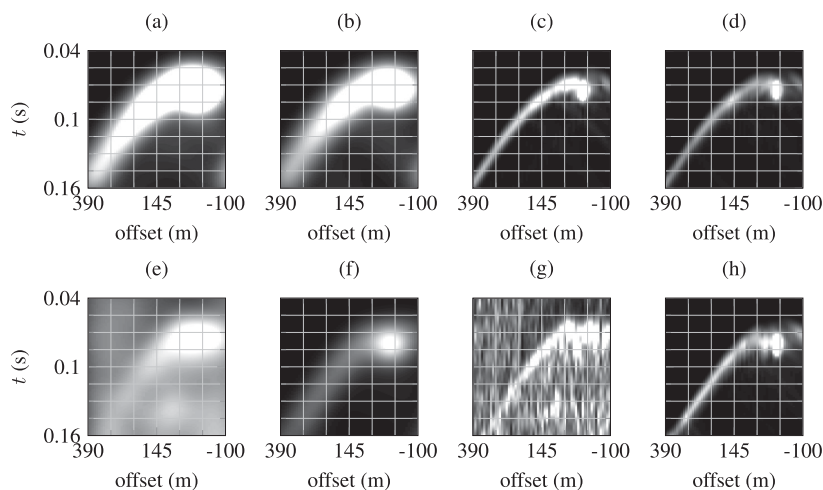
Figure 16. Shot gathers corresponding to the cross-well experiment in Figs 15 and 18 for a source at (205, 199) m. (a) Synthetic data using interpolated well-log models. (b) Observed field data. (c) Initially modelled data for synthetic and field data inversion. (d) Modelled data after synthetic data multi-objective inversion, to be compared to (a). (e) Modelled data after field data inversion, to be compared to (b). (f) Same as (e), but after single-objective J_{ls} inversion.
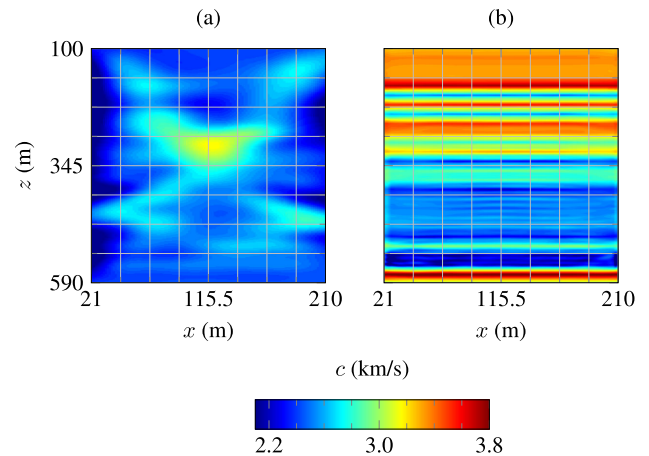
the modelled bumpy data after the inversion, plotted in Figs 17(b) and (d), match the observed bumpy data.

5.1.2 Field data inversion

Under the acoustic approximation, we assumed that the recorded x-component data generated by x-component sources are not too different from acoustic pressure data generated by explosive sources. We band-limited the observed data to the range between 80 and 120 Hz, such that the low frequencies are removed. Attenuation is not taken into account, but during pre-processing, an amplitude-versus-offset correction was applied to the observed data using the initially modelled data, following Brenders & Pratt (2007). The observed shot gather for a source at (205, 199) is shown in Fig. 16(b). Both single- and multi-objective inversions in this case are per-

Figure 18. A cross-well example with field data and well-logs plotted in Figs 15(a) and (b). The output ρ models are not shown. (a) Reconstructed c model from the field data using the single-objective J_{ls} inversion. (b) Same as (a), but using the multi-objective inversion strategy.

formed in two stages. During the first stage, the subsurface models are allowed to vary only with depth. Then, the output is used as a starting model to estimate the final results that vary in both the x- and z-coordinates.

Single-objective J_{ls} inversion outputs the velocity model plotted in Fig. 18(a) that varies both in the x- and z-coordinates. The corresponding modelled shot gather in Fig. 16(f) is cycle skipped compared to the observed shot gather in Fig. 16(b): note the arrivals around the red square. The multi-objective inversion results in the final velocity model plotted in Fig. 18(b) that also varies both in the x- and z-coordinates. The modelled shot gather after inversion, plotted in Fig. 16(e), matches the observed shot gather in Fig. 16(b). The modelled bumpy data, plotted in Figs 17(f) and (h), match the observed bumpy data in Figs 17(e) and (g) after the inversion. Finally, Fig. 19 displays vertical cross-sections of the synthetic- and field-data multi-objective inversion results.
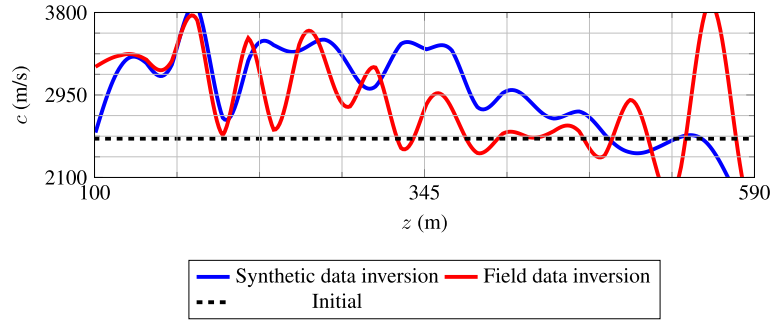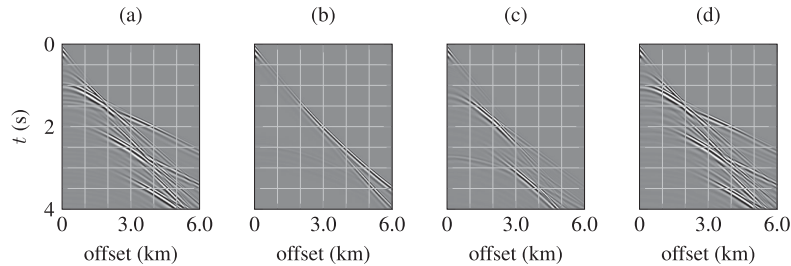

Figure 17. Bumpy shot gathers corresponding to the cross-well experiment in Figs 15 and 18 for a source at (205, 199) m. (a) Observed and (b) modelled gathers after synthetic data inversion using $\sigma_t/\tau_d = 1.5$ and $\sigma_r/\lambda_d = 1.0$. (c) Observed and (d) modelled gathers after synthetic data inversion using $\sigma_t/\tau_d = 0.5$ and $\sigma_r = 0$. (e) Observed and (f) modelled gathers after field data inversion using $\sigma_t/\tau_d = 1.5$ and $\sigma_r/\lambda_d = 1.0$. (g) Observed and (h) modelled gathers after field data inversion using $\sigma_t/\tau_d = 0.5$ and $\sigma_r = 0$.

Figure 19. Output models for the cross-well experiment. For the synthetic data inversion, the output in Fig. 15(c) varies only with depth and is plotted in blue. A cross-section at $x = 110$ m, obtained from the field data and corresponding to the reconstructed model in Fig. 18(b), is plotted in red.
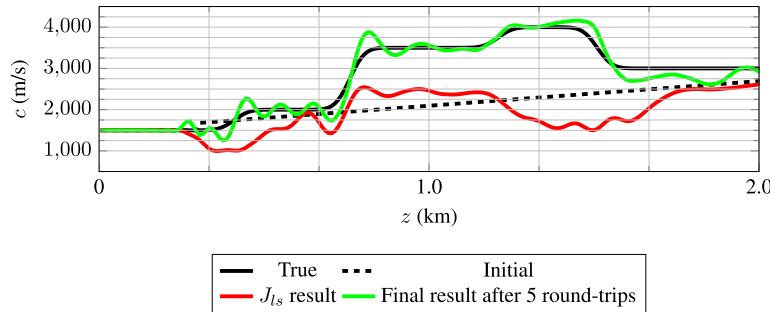
Figure 20. Plot corresponding to the five-layer example. The actual (solid black) and initial (dashed black) models are plotted. Single-objective J_{ls} inversion result is plotted in red. The multi-objective inversion scheme, depicted in Fig. 12, results in a model plotted in green.

Figure 21. The shot gathers corresponding to the results in Fig. 20. (a) Observed gather. (b) Initially modelled gather. (c) Modelled gather after single-objective J_{ls} inversion. (d) Modelled gather after multi-objective inversion.

5.2 Five-layer example

We now numerically test the applicability of the multi-objective strategy when:

(i) the data are acquired at sufficiently long offsets to contain refracted waves and

(ii) the data contain many arrivals including free-surface reflection as well as refraction multiples.

A somewhat more realistic five-layer Earth model of 6000 m width and 2500 m depth is assumed and the velocity varies with depth as depicted in Fig. 20. We placed an evenly spaced horizontal array of 200 receivers at a depth of 10 m. A source is placed at (0, 10) such that the maximum offset in the data is 6000 m. We generated 'observed' data using a fourth-order minimum-phase Butterworth source wavelet of bandwidth 5–10 Hz. The 'observed' shot gather, plotted in Fig. 21(a), contains refractions due to high-velocity contrasts at the interfaces just below 750 and 1200 m. We choose a free-surface boundary condition so that free-surface multiples are present in the data. The initial velocity model, plotted in Fig. 20,

has an incorrect background velocity. The modelled data using the initial velocity model are plotted in Fig. 21(b).

We now perform the multi-objective inversion depicted in Fig. 12. Each round-trip consists of two different optimization stages. During the first stage, we used the blurring parameters as in Table 2 with time windowing to fit only the early refracted arrivals at offsets larger than 2000 m. The inversion model grid spacing during these optimizations is 50 m and a Gaussian with standard deviation of 210 m is used for gradient smoothing. In the next stage, the optimizations minimize the error in the entire shot gather without windowing and use the blurring parameters given in Table 2. The model spatial sampling is 30 m. The standard deviation of the Gaussian used for gradient smoothing is 60 m. The modelled data after five round-trips, plotted in Fig. 21(d), match the synthetic observed data. The corresponding output velocity model is plotted in Fig. 20.

We now perform the single-objective J_{ls} inversion also in two stages that are described above. Single-objective J_{ls} inversion fails to reconstruct the background velocity of the model, as plotted in Fig. 20. Also, the modelled data at the last iteration, plotted in Fig. 21(c), do not fit the observed data.
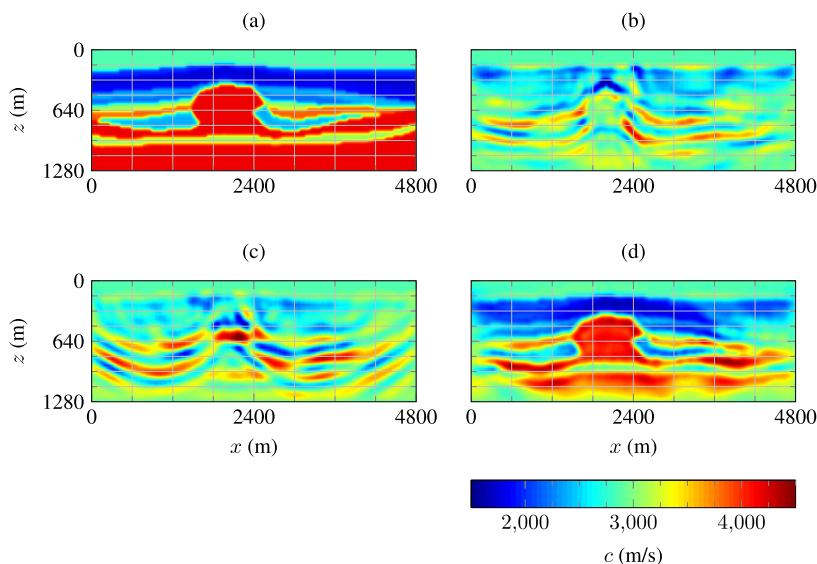
Figure 22. A 2-D reflection experiment on a complex model with a high-velocity structure. The initial model for inversion is homogeneous with $c = 3000 \, \mathrm{m \, s^{-1}}$. (a) Actual velocity model. (b) Single-objective J_{ls} inversion result. (c) Single-objective J_b inversion result using $\sigma_t/\tau_d = 1$ and $\sigma_r/\lambda_d = 0.5$. (d) Result from the multi-objective inversion strategy, outlined in Fig. 12, after 10 inversion round-trips.

5.3 Complex 2-D model with reflection data

Fig. 22(a) depicts an assumed Earth model of 4800 m width and 1280 m depth with a high-velocity structure. We placed an evenly spaced horizontal array of 60 sources and 100 receivers at a depth of 20 m. We used a 5–12-Hz fourth-order minimum-phase Butterworth source wavelet. The 'observed' data generated for the actual model do not contain any refracted arrival due to the negative velocity anomaly at shallow depths. The initial velocity model for inversion is homogeneous with $c = 3000 \, \mathrm{m \, s^{-1}}$. The background velocity of the model has to be reconstructed using only the reflected arrivals in the data. In order to reduce high-wavenumber artefacts, we apply Gaussian smoothing to the gradient at each iteration with a standard deviation of 32 m. The single-objective J_{ls} inversion results in a model with an incorrect background velocity, plotted in Fig. 22(b). We started again from the homogeneous model but now with the bump functional, setting $\sigma_t/\tau_d = 1$ and $\sigma_r/\lambda_d = 0.5$. As expected, the single-objective bump functional inversion results in a model that does not have the correct background velocity, as can be seen in Fig. 22(c). Finally, we performed 10 round-trips using the multi-objective inversion strategy, outlined in Fig. 12. The multi-objective inversion uses the blurring parameters as in Table 2. The resulting velocity model is plotted in Fig. 22(d) and resembles the actual velocity model fairly well.

(ii) blurring increases the size of the basin of attraction that corresponds to the functional and hence its global-convergence robustness;

(iii) the bump-functional inversion suffers from the fact that the model that matches given bumpy data tends to be highly non-unique.

Single-objective bump-functional inversion can produce acceptable results while fitting transmitted arrivals. Whereas in the case of reflected arrivals, the non-uniqueness prevents the bump functional to update the background velocity of the model. Therefore, in order to reach the global minimum corresponding to the least-squares objective, we proposed a multi-objective inversion scheme that uses the bump functional as an auxiliary functional. We demonstrated the potential of the bump functional to pull the trapped solution out of the least-squares local minimum whenever necessary. Finally, we have tested the applicability of the multi-objective inversion scheme using realistic numerical examples as well as cross-well field data. In all the cases, the scheme found the model that well explains the observed data and hence corresponds to the global minimum of the least-squares functional, even in the absence of low frequencies in the data.

6 CONCLUSIONS

We have formulated a data-domain functional that matches the observed and the modelled data in a simplified form. It can be viewed as a generalized envelope-based misfit. The simplification results in bumpy data, obtained by taking the absolute value of the data and subsequent smoothing or blurring with a Gaussian. Using numerical examples involving transmission or reflection data, we illustrated the following characteristics of the bump functional:

(i) the functional is less sensitive to cycle skipping and does not rely on the low frequencies present in the data;

REFERENCES

Bozdağ, E., Trampert, J. & Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements, *Geophys. J. Int.,* **185**(2), 845–870.

Brenders, A. & Pratt, R., 2007. Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model, *Geophys. J. Int.,* **168**(1), 133–151.

Bunks, C., Saleck, F.M., Zaleski, S. & Chavent, G., 1995. Multiscale seismic waveform inversion, *Geophysics,* **60**(5), 1457–1473.

Chauris, H., Donno, D. & Calandra, H., 2012. Velocity estimation with the normalized integration method, in *74th EAGE Conference & Exhibition incorporating SPE EUROPEC,* The European Association of Geoscientists and Engineers.

Chi, B., Dong, L. & Liu, Y., 2014. Full waveform inversion method using envelope objective function without low frequency data, *J. Appl. Geophys.,* **109**, 36–46.

Datta, D., 2015. Estimating starting models for full waveform inversion using a global optimization method, in *77th EAGE Conference and Exhibition 2015,* The European Association of Geoscientists and Engineers.

Donno, D., Chauris, H. & Calandra, H., 2013. Estimating the background velocity model with the normalized integration method, in *75th EAGE Conference & Exhibition incorporating SPE EUROPEC,* The European Association of Geoscientists and Engineers.

Engquist, B. & Froese, B.D., 2013. Application of the Wasserstein metric to seismic signals, arXiv:1311.4581.

Fei, T.W., Luo, Y., Qin, F. & Kelamis, P.G., 2012. Full waveform inversion without low frequencies: a synthetic study, in *SEG Technical Program Expanded Abstracts,* Society of Exploration Geophysicists.

Fichtner, A., 2010. *Full Seismic Waveform Modelling and Inversion,* Springer.

Gao, Z., Pan, Z., Gao, J. & Zhang, X., 2014. Building an initial model for full waveform inversion using a global optimization scheme, in *SEG Technical Program Expanded Abstracts,* Society of Exploration Geophysicists.

Gauthier, O., Virieux, J. & Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveforms: numerical results, *Geophysics,* **51**(7), 1387–1403.

Luo, J. & Wu, R.-S., 2015. Seismic envelope inversion: reduction of local minima and noise resistance, *Geophys. Prospect.,* **63**(3), 597–614.

Luo, S. & Sava, P., 2011. A deconvolution-based objective function for wave-equation inversion, in *SEG Technical Program Expanded Abstracts,* Society of Exploration Geophysicists.

Luo, Y. & Schuster, G., 1991. Wave-equation traveltime inversion, *Geophysics,* **56**(3), 645–653.

Ma, Y., Hale, D., Gong, B. & Meng, Z., 2012. Image-guided sparse-model full waveform inversion, *Geophysics,* **77**(4), R189–R198.

Mora, P., 1988. Elastic wave-field inversion of reflection and transmission data, *Geophysics,* **53**(6), 750–759.

Mora, P., 1989. Inversion = migration + tomography, *Geophysics,* **54**(12), 1575–1586.

Mulder, W.A. & Plessix, R.-E., 2008. Exploring some issues in acoustic full waveform inversion, *Geophys. Prospect.,* **56**(6), 827–841.

Plessix, R.-E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.,* **167**, 495–503.

Plessix, R.-É. & Mulder, W.A., 2004. Frequency-domain finite-difference amplitude-preserving migration, *Geophys. J. Int.,* **157**(3), 975–987.

Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain, Part 1: theory and verification in a physical scale model, *Geophysics,* **64**(3), 888–901.

Sambridge, M. & Drijkoningen, G., 1992. Genetic algorithms in seismic waveform inversion, *Geophys. J. Int.,* **109**(2), 323–342.

Schuster, G.T., 2015. *Seismic Imaging,* SEG Press.

Sen, M.K. & Stoffa, P.L., 1991. Nonlinear one-dimensional seismic waveform inversion using simulated annealing, *Geophysics,* **56**(10), 1624–1638.

Shin, C. & Cha, Y.H., 2008. Waveform inversion in the Laplace domain, *Geophys. J. Int.,* **173**(3), 922–931.

Shin, C. & Cha, Y.H., 2009. Waveform inversion in the Laplace-Fourier domain, *Geophys. J. Int.,* **177**(3), 1067–1079.

Shin, C. & Min, D.-J., 2006. Waveform inversion using a logarithmic wavefield, *Geophysics,* **71**(3), R31–R42.

Snieder, R., Xie, M., Pica, A. & Tarantola, A., 1989. Retrieving both the impedance contrast and background velocity: a global strategy for the seismic reflection problem, *Geophysics,* **54**(8), 991–1000.

Stoffa, P.L. & Sen, M.K., 1991. Nonlinear multiparameter optimization using genetic algorithms: inversion of plane-wave seismograms, *Geophysics,* **56**(11), 1794–1810.

Sun, D. & Symes, W.W., 2012. Waveform inversion via nonlinear differential semblance optimization, in *SEG Technical Program Expanded Abstracts,* Society of Exploration Geophysicists.

Symes, W., 2008. Migration velocity analysis and waveform inversion, *Geophys. Prospect.,* **56**(6), 765–790.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics,* **49**(8), 1259–1266.

van Leeuwen, T., 2010. Correlation-based seismic velocity inversion, *PhD thesis,* Delft University of Technology.

van Leeuwen, T. & Mulder, W.A., 2010. A correlation-based misfit criterion for wave-equation traveltime tomography, *Geophys. J. Int.,* **182**(3), 1383–1394.

Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics,* **74**(6), WCC1–WCC26.

Warner, M. & Guasch, L., 2014. Adaptive waveform inversion: Theory, in *SEG Technical Program Expanded Abstracts,* Society of Exploration Geophysicists.

Wu, R.-S., Luo, J. & Wu, B., 2014. Seismic envelope inversion and modulation signal model, *Geophysics,* **79**(3), WA13–WA24.

Zhang, Y. & Wang, D., 2009. Traveltime information-based wave-equation inversion, *Geophysics,* **74**(6), WCC27–WCC36.

APPENDIX: ADJOINT SOURCE FUNCTIONS FOR THE BUMP FUNCTIONAL

The gradient of the bump functional with respect to the medium parameters, c and ρ, is computed by correlating the forward-propagated source wavefield with the adjoint wavefield at each point in the subsurface. This is similar to the gradient computation for the conventional least-squares inversion, where the adjoint wavefield is generated by injecting the adjoint source functions from the receiver positions (Tarantola 1984; Plessix 2006; Fichtner 2010). The adjoint source functions for the least-squares functional are the difference between the modelled and the observed data. In order to compute the adjoint source functions for the bump functional, the following steps are performed in order:

(i) the difference between the modelled and the observed bumpy data is calculated,

(ii) time and/or receiver blurring is applied to the difference, based on σ_t and σ_r, and

(iii) the result is multiplied with the *stabilized* sign of the modelled data in the time domain.

The adjoint source functions for the bump functional are given by $\nabla_p J_b$, where ∇_p denotes the derivative with respect to the modelled data p. Using the chain rule, we write:

$$\nabla_p J_b = \nabla_{p_a} J_b \, \nabla_p p_a,$$
$$= \nabla_{p_a} J_b \underbrace{p \, [p^2 + \epsilon^2]^{-\frac{1}{2}}}_{\text{sgn}(p), \text{ when } \epsilon \text{ is } 0}, \tag{A1}$$

where ∇_{p_a} denotes the derivative with respect to the absolute-valued-modelled data p_a and sgn is the sign function. In order to derive an expression for $\nabla_{p_a} J_b$, we consider the bump functional in frequency–wavenumber domain. In such a domain, time and/or receiver blurring corresponds to a simple multiplication with the blurring function $\tilde{b}$. Without loss of generality, we consider the case of a regular horizontal array of receivers that record waves originating from a given source location. Using Parseval's theorem, we rewrite the expression of the bump functional in eq. (4) by Fourier transforming the bumpy data from time t and horizontal receiver coordinate x_r to frequency f and horizontal wavenumber k_r as

$$J_b = \frac{1}{2} \sum_r \sum_t [p_b(x_r, t) - q_b(x_r, t)]^2,$$

$$= \frac{1}{8\pi^2} \sum_r \sum_f |\tilde{p}_b(k_r, f) - \tilde{q}_b(k_r, f)|^2, \tag{A2}$$

where we used the fact that p_b and q_b are real valued. Here, $\tilde{p}_b$ and $\tilde{q}_b$ denote the modelled and the observed bumpy data in the f–k_r domain, respectively. In the f–k_r domain, the bumpy data are obtained by multiplying the absolute-valued data, $\tilde{p}_a$ and $\tilde{q}_a$, with the blurring function $\tilde{b}$, so we have

$$J_b = \frac{1}{8\pi^2} \sum_r \sum_f |\tilde{b}(f, \sigma_t)\tilde{b}(k_r, \sigma_r)[\tilde{p}_a(k_r, f) - \tilde{q}_a(k_r, f)]|^2. \tag{A3}$$

We rewrite the above equation using the real and imaginary parts of the absolute-valued data as

$$J_b = \frac{1}{8\pi^2} \sum_r \sum_f \tilde{b}^2(f, \sigma_t)\tilde{b}^2(k_r, \sigma_r)$$

$$\times \{[\Re(\tilde{p}_a) - \Re(\tilde{q}_a)]^2 + [\Im(\tilde{p}_a) - \Im(\tilde{q}_a)]^2\}, \tag{A4}$$

where we used the fact that the blurring function $\tilde{b}$ is real valued. We now differentiate the above equation with respect to the real and imaginary parts of the modelled-absolute-valued data $\tilde{p}_a$ to obtain

$$\nabla_{\tilde{p}_a} J_b = \nabla_{\Re(\tilde{p}_a)} J_b + \iota \nabla_{\Im(\tilde{p}_a)} J_b$$

$$= \frac{1}{4\pi^2} \tilde{b}^2(f, \sigma_t)\tilde{b}^2(k_r, \sigma_r)[\tilde{p}_a(k_r, f) - \tilde{q}_a(k_r, f)],$$

$$= \frac{1}{4\pi^2} \tilde{b}(f, \sigma_t)\tilde{b}(k_r, \sigma_r)[\tilde{p}_b(k_r, f) - \tilde{q}_b(k_r, f)]. \tag{A5}$$

Using the chain rule, we can write the derivative of the bump functional with respect to modelled-absolute-valued data in the time domain as

$$\nabla_{p_a} J_b = \sum_r \sum_f \nabla_{\tilde{p}_a} J_b \nabla_{p_a} \tilde{p}_a,$$

$$= b(x_r, \sigma_r) *_r \{b(t, \sigma_t) *_t [p_b(x_r, t) - q_b(x_r, t)]\}. \tag{A6}$$

After substituting $\nabla_{p_a} J_b$ in eq. (A1), we obtain the final expression to compute the adjoint source functions for the bump functional in the time domain:

$$\nabla_p J_b = \{b(x_r, \sigma_r) *_r \{b(t, \sigma_t) *_t [p_b - q_b]\}\} \, p \, [p^2 + \epsilon^2]^{-\frac{1}{2}}. \tag{A7}$$